AD_____

Award Number: DAMD17-94-J-4345

TITLE: Evaluation of Digital Mammography Display

PRINCIPAL INVESTIGATOR: Etta Pisano, M.D.

CONTRACTING ORGANIZATION: University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599

REPORT DATE: September 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release
distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1999	Final (01 Sep 94 - 31 Aug 99)

4. TITLE AND SUBTITLE
Evaluation of Digital Mammography Display

5. FUNDING NUMBERS
DAMD17-94-J-4345

6. AUTHOR(S)
Etta Pisano, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599

e-mail:
etpisano@med.unc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The purpose of this research is to experimentally determine the diagnostic accuracy and clinical acceptability of digitally acquired mammograms displayed on soft copy display compared to laser printed hard copy. We have conducted observer studies both under laboratory conditions and under simulated conditions. We have used computer generated lesions and we have used real clinical mammograms to evaluate different image processing techniques. Our preliminary results indicate that choice image processing method is a function of the type of lesion and machine type.
We have developed a mammography workstation that is easy to use and fast. The observer studies that will determine the diagnostic accuracy and acceptability of the digital mammograms and the soft copy display are presently under way and the results will be known by the end of the year, 1999.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
131

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

_✓___ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Ethel Rios 9/30/99

PI - Signature Date

TABLE OF CONTENTS

Introduction

a) Nature of the problem (abbreviated from original text)

A full-field digital mammography system has been developed by Fischer Medical systems in collaboration with the University of Toronto. This scanning slot digital mammography system provides 50um, 12-bit pixels with inherently better contrast than that of conventional mammogram. The advent of digitally acquired mammograms offers the possibility of further improvements in early breast cancer detection. Specifically, digital acquisition systems decouple the process of x-ray photon detection from image display by using a primary detector that directly quantifies transmitted photons. This allows digital systems to be more efficient in utilization of radiation dose. Digital systems also allow a wide dynamic range so that a wider range of tissue contrast can be appreciated. Subtle contrast differences can be amplified and the distinction between benign and malignant might be increased. The new scanning slot digital mammography system has the further advantage of reduced scatter compared with both conventional and phosphor plate technologies. Furthermore, digital systems have the capacity to bring revolutionary advantages to breast cancer detection and management: 1) image processing for increased lesion conspicuity; 2) computer-aided diagnosis for enhanced radiologic interpretation; 3) teleradiology, or image transmission, as a means of bringing world-class expertise to community hospitals and remote areas; 4) improved image access and communication through digital image archiving and transmission; and 5) dynamic, or "real time" imaging for use during biopsy and localization procedures.

b) Purpose of this research

The purpose of this study is to determine experimentally the diagnostic accuracy and interpretation speed of the available display methods.

c) Methods of approach

We propose to conduct an ROC study involving the best available display methods, one representative of a film based display, and one using the best available state-of-the-art electronic workstation.

Body

Accomplishment 1.

To achieve the goals of this research, we used full field digitally acquired mammograms. Availablity of the clinical digital units were delayed because of detector upgrades and manufacturing problems. However, our Fischer unit was installed at UNC Hospitals in April of 1997. In Jan. 1998 Fischer upgraded the system with a new detector that improved resolution and reliability of the system. We have completed acquisition of digital mammograms with a total of than 300 clinical mammograms.

Accomplishment 2.

During the first part of this grant, a number of changes in the state-of-the-art of monitor technology occured, a) High brightness/resolution monitors, although commercially available, were not as readily available as once promised. There continue to be manufacturing problems in quality assurance and meeting performance specifications. We have evaluated a number of different brands in our laboratory and with collaboration of Dr. Hans Rhoerig at Univ. of Arizona and Dr. Harwig Blume at Philips Medical. As a result of these extensive evaluations, we purchased two DataRay and two Orwin monitors. To achieve the maximum displayable grey -levels, we installed the electronics from Dome (10 bits grey level). We have developed interactive software that provides a viable mammography workstation. This software has been implemented and is being used for the observer study now under way.

Accomplishment 3.

We have conducted a "preference" observer study to evaluate eight different methods of image processing for display of digitally acquired mammograms. The complete paper is included as Appendix A.1.

Mammograms were acquired on three different full-field digital mammographic systems and were displayed on laser printed film. The eight different techniques are as follows: 1) hand intensity windowing, 2) Peripheral equalization followed by hand intensity windowing, 3) unsharpmasking followed by hand intensity windowing, 4) Contrast Limited Adaptive Histogram 5) Mixture modeling based intensity windowing, 6) Hhistogram based intensity windowing, 7) MUSICA and 8) TREX proprietary processing method.

In summary, a total of 8 processed images for each of the 28 cases were compared to film screen images by 12 radiologists. The total number of images viewed per radiologist was 497, with the group as a whole reviewing 5964 images. The cases contained a total of 65 lesions, 29 that were pathologically proven and 36 that were presumed benign. Since there were two scores for each lesion (cc and mlo) for each algorithm for the diagnostic task, and an additional score for each view for each algorithm for the screening task, the total number of scores requested per radiologist was 1439, and 17268 scores requested from the whole group.

Results: Primary Analysis: Diagnostic Mammography Task

There was a strongly statistically significant relationship between lesion type and image processing algorithm preference. ($p=0.0019$) That is to say, radiologists preferred different algorithms for each of the two tasks, that is, mass characterization and calcification characterization.

a) For the diagnostic evaluation of masses (including masses with calcifications), the printed digital mammogram was preferred to the film screen radiograph for all eight processing algorithms. The mean scores ranged from +0.28 down to +0.01. Unsharp Masking (UM) received the highest mean score and MMIW received the lowest mean score. Only UM was rated as significantly better than the film screen mammogram for mass evaluation (alpha=.01/16=.000625). For the mass characterization task, pairwise comparisons revealed several strongly significant differences ($p \leq 0.000714285$) in radiologist preferences. Unsharp Masking was preferred to MIW, CLAHE and MMIW. Both Manual Intensity Windowing (MIW) and CLAHE were preferred to Mixture Model Intensity Windowing (MMIW).

b)For the diagnostic evaluation of calcifications, the film screen radiograph was preferred to the printed digital mammogram for all eight processing algorithms. The mean scores ranged from –0.09 down to –0.71. HIW received the highest mean score and PE received the lowest mean score. PE, CLAHE, MIW and MUSICA were rated as significantly worse than the film screen mammogram for the evaluation of calcifications (alpha=.01/16=.000625). For the calcification characterization task, pairwise comparisons revealed several strongly significant differences ($p \leq 0.00714285$) in radiologist preferences. All algorithms were preferred over Peripheral Equalization. Trex processing was preferred to both MIW and CLAHE.

c) Summary:The test of interaction between processing algorithm and lesion type was highly significant ($p=0.0019$). Although the mean score is negative for calcifications and positive for masses for each algorithm, the difference between the mean scores for calcifications and masses varies across algorithms. Given the significance of the interaction test, the main effect tests for algorithm and lesion type are irrelevant.

Resuls: Secondary Analysis: Overall Screening Task:

With respect to screening, the film screen radiograph was preferred to the printed digital mammogram for all eight processing algorithms, with mean scores ranging from –0.26 (Trex) down to –1.25 (MMIW). Each test of the mean score equal to 0 was evaluated at the .01/8=.00125 level. Algorithms with p-values<.00125 were HIW, MIW, CLAHE, UM, PE and MMIW. Since this is an exploratory analysis, p-values may only be interpreted as descriptive statistics, and not as tests of significance.

This study is limited by the fact that it was a preference study and not a quantitative measure of how well the radiologists performed. Radiologists gave their opinions on which images would improve their performance. Certainly they made educated guesses, but a performance study would have been better at determining how

mammographic interpretation would be affected by image processing. This study is a good first step, however.

Accomplishment 4.

Multi-center Clinical Evaluation of Digital Mammography

This project is a multi-center clinical trial designed to determine whether digital mammography can improve the detection and characterization of breast lesions in the population of patients presenting for problem-solving mammography. Through this study, 380 consecutive eligible women who presented for problem-solving mammography at 8 mammographic centers in the United States and Canada who underwent breast biopsy, and a random sample of those who did not undergo biopsy, were enrolled in a trial where they had digital mammography. These studies and the film-screen mammograms of the same patients are read in a controlled experimental reading study involving 18- radiologists using a 5 point scale suitable for ROC analysis. For this project, we will compare the ROC curves for the radiologist's interpretations of the film-screen mammograms, with and without additional views and sonograms, to the ROC curves for their interpretations of the default digital mammograms and an imaged processed version of the digital mammograms displayed on film.

Our hypothesis is that digital mammography will improve radiologist's performance in diagnosing breast cancer compared to their performance using film-screen mammography in the population of patients presenting for problem-solving mammography, as measured by the area under the ROC curve. This observer study is in process with 16 of the 18 observers completed. The target date for completion of this study is September 30, 1999, with ROC analysis completed by the middle of November, 1999. Upon completion the final analysis will be reported to the Army.

Accomplishment 5.

Comparison of film display to softcopy display:

The purpose of this study is to compare the diagnostic accuracy and reading times of mammography film readings to video monitor readings including diagnostic accuracy and interpretation time. The study is to interpret approximately 132 mammograms, half on a video display and half on film screen on a lightbox. Each case will consist of a current 4 view mammogram, and a previous (approximately 1 year old mammogram). The current mammogram will be a digital mammogram; the previous one will be a film mammogram.

a) Subject population: All of the women imaged came from the diagnostic or problem-solving mammography population at various institutions around the US and Canada. These women had palpable lumps, discharges, and abnormal screening mammograms as their purpose for seeking diagnostic mammography. Some underwent breast biopsy. Some of them were recommended to undergo only annual or 6 month follow-up mammography.

b) Methods

The reader is shown the standard mammographic images that were done at the time of their diagnostic mammography visit. The reader is asked to perform the reading quickly but without errors, as they would in the clinic. This is so that we can measure reading times using film and video displays to see if they are similar, or different. The reader will first read and report on the films. This portion will be timed so that we can record how long a standard reading and dictation take. The reader will then review the films and describe the findings to the research assistant. This part is not timed. The RA will ask specific questions about the lesions identified and will fill out forms noting the responses. The types of lesions in this study set are the same ones in everyday practice, i.e. masses (with or without associated calcifications), calcifications, architectural distortions and asymmetric densities. The reader is asked to grade every lesion detected using two different scales, as follows. Note that we are NOT using the BIRADS classification scheme because it is not suitable to the task of this research.

1) The finding is definitely not malignant.

2) The finding is probably not malignant.
3) The finding is possibly malignant.
4) The finding is probably malignant.
5) The finding is definitely malignant.

What would you recommend for this finding?

1) No further work-up. Routine follow-up only.
2) No further work-up. Six month follow-up only.
3) Further work-up with additional mammographic views, and/or ultrasound.
4) Further work-up with either percutaneous or open surgical biopsy.

In addition, the following information will be provided for all types of lesions that are identified: Side (Right or Left), O'clock location (1-12, Axillary tail or straight back from the nipple). If it is seen with certainty on only one view, indicate on which view it is seen (mlo or cc) and where in the plane of that view it is located (for mlo, superior, mid or inferior and for cc, medial, mid or lateral),AP location (anterior, central, posterior). For every clinically relevant ARCHITECTURAL DISTORTION seen, there is no additional information needed beyond that needed for all lesions. Architectural distortions associated with masses do not have to be recorded separately. For every clinically relevant ASYMMETRIC DENSITY seen, there is no additional information needed beyond that needed for all lesions. Again, asymmetric densities associated with masses do not have to be recorded separately. Clustered calcifications associated with asymmetric densities should be recorded primarily as clustered calcifications and the asymmetry should be noted as a separate finding with the calcifications as an associated feature.
This final study is in the final stages of preparation and is scheduled to begin the end of September, 1999 and to be completed by the middle of November, 1999. Analysis of the results will be available by the end of December, 1999. The results will be forwarded to the Army upon completion..

NOTE: "Accomplishments" 4 and 5 of the experimental work is presently in process. The observers studies are behind schedule because of delays in acquiring full field digital mammograms from the various participating sites, formatting problems in applying the different image processing algorithms to the different manufactures and scheduling observers from participating institutions. These studies will be completed with analysis by the end of December, 1999.

Key Research Accomplishments

- Acquisition of data base of full field digital mammograms (partially supported by this grant and under the auspices of the International Digital Mammography Development Group).
- Development of dual screen soft copy mammography workstation.
- Evaluation of multiple methods of image processing for both hard and soft copy display.
- Comparison of diagnostic accuracy of soft copy to hard copy display.

Reportable Outcomes

Manuscripts, abstracts, presentations.

A. Publications/Manuscripts:

1. Pisano ED, Chandramouli J, Hemminger BM, DeLuca M, Glueck D, Johnston RE, Muller K, Braeuning MP, Pizer S. Does intensity windowing improve the detection of simulated calcifications in dense mammograms? Journal of Digital Imaging 1997;10(2):79-84.
2. Pisano ED, Chandramouli J, Hemminger BM, Johnston RE, Muller K, Pizer S. The effect of intensity windowing as an image processing tool in the detection of simulated masses embedded in digitized mammograms. Journal of Digital Imaging 1997;10(4):174-182.
3. Hemminger BM, Dillon A, Johnston RE, Muller K, Pisano ED, Deluca M. Evaluation of the effect of display luminance on the feature detection of simulated masses in mammograms. SPIE Medical Imaging 1997;3036:12.
4. Pisano ED, Zong S, Hemminger BM, DeLuca M, Johnston RE, Muller K, Braeuning MP, Pizer S. Contrast Limited Adaptive Histogram Equalization Image Processing to Improve the Detection of Simulated Spiculations in Dense Mammograms. Journal of Digital Imaging. 1998; 11(4): 193-200.
5. Pisano ED, Yaffe M. Digital mammography. Breast Disease 1998;10(3,4):127-136.
6. Pisano ED, Yaffe M. Digital mammography. Contemporary Diagnostic Radiology1998;21(15):1-6.
7. Pisano ED. Initial clinical experience with full field digital mammography. Proceedings of the Fourth International Workshop on Digital Mammography 1998;13:391-394.
8. Aylward SR, Hemminger BM, Pisano ED. Mixture modeling for digital mammogram display and analysis. Proceedings of the Fourth International Workshop on Digital Mammography 1998;13:305-312.
9. Pisano ED, Yaffe M, Hemminger BM, Hendrick E, Niklason L, Maidment A, Kimme-Smith C, Feig S, Sickles E. Current Status of Full-Field Digital Mammography. Submitted to Radiology.
10. Pisano ED, Aylward S, Barbour P, Braeuning M.P, Brown ME, Chakraborty D, Cole E, Conant E, Eagle E, Fajardo LL, Feig S, Harrison J, Hemminger BM, R. Johnston RE, Jong R, Kennedy R, Kopans D, Kornguth P, Maidment A, Major S, McLelland R, Moore R, Muller K, Niklason L, Nishikawa R, Pizer SM, Plewes DB, Rosen E, Poyet C, Seaton K, Soo MS, Shumak R, Stahpit S, Staiger M, Vermont A, Walsh R, Williams MB, Williford M, Yaffe M, and Zong Z. Radiologist Preferences for Imaging Processing Algorithm for different clinical tasks for digital mammography display. To be submitted to Radiology.
11. Pisano ED, Aylward S, Barbour P, Braeuning M.P, Brown ME, Chakraborty D, Cole E, Conant E, Eagle E, Fajardo LL, Feig S, Harrison J, Hemminger BM, R. Johnston RE, Jong R, Kennedy R, Kopans D, Kornguth P, Maidment A, Major S, McLelland R, Moore R, Muller K, Niklason L, Nishikawa R, Pizer SM, Plewes DB, Rosen E, Poyet C, Seaton K, Soo MS, Shumak R, Stahpit S, Staiger M, Vermont A, Walsh R, Williams MB, Williford M, Yaffe M, and Zong Z. Pictorial Essay on the Use of Different

B. Presentations/Abstracts.

1. Pisano ED, Aylward S, Barbou Visualization for Pre-operative Diagnostic Evaluation and Surgical Planning. Inforad Exhibitor (RSNA, 1995)

2. Hemminger B, Pisano ED, Johnston RE, et al. Mammographic Image Display Using a Workstation. Inforad Exhibitor (RSNA, 1996)

3. Pisano ED, Hemminger BM, Johnston RE, Muller K. A Prototype Digital Mammography Workstation. Department of Defense Era of Hope Conference on Breast Cancer Research. Washington, DC. November 1997.

4. Hemminger B, Pisano ED, Johnston RE, et al. Workstation for Digital Mammography. Inforad Exhibitor (RSNA, 1997)

5. Hemminger BM, Pisano ED, Stahpit S, Johnston RE. Demonstration of Softcopy Display System for Digital Mammography. Radiologic Society of North America Meeting. Chicago, IL. November 29-December 4, 1998. Inforad Exhibitor (RSNA, 1998). P, Braeuning M.P, Brown ME, Chakraborty D, Cole E, Conant E, Eagle E, Fajardo LL, Feig S, Harrison J, Hemminger BM, R. Johnston RE, Jong R, Kennedy R, Kopans D, Kornguth P, Maidment A, Major S, McLelland R, Moore R, Muller K, Niklason L, Nishikawa R, Pizer SM, Plewes DB, Rosen E, Poyet C, Seaton K, Soo MS, Shumak R, Stahpit S, Staiger M, Vermont A, Walsh R, Williams MB, Williford M, Yaffe M, and Zong Z. Comparison of the Acceptability and Performance of Image Processing Algorithms in Visualizing Known Lesions in Digital Mammograms. Radiologic Society of North America Meeting. Chicago, IL. November 29-December 4, 1998. Awarded Certificate of Merit and invited for publication in Radiographics.

6. Pisano ED. Image Processing in Digital Mammography: Dynamic Intensity Windowing as a tool to improve mass detection in digitized mammograms. National Digital Mammography Development Group Meeting. Philadelphia, Pa. June 13, 1995.

7. Pisano ED. Digital Mammography. University of California Post-graduate Course in Breast Imaging. McLean, VA, September 17, 1995.

8. Pisano ED, Chandramouli J, Hemminger B, Johnston RE, Pizer S, Muller K. Utility of Intensity Windowing in Improved Detection of Simulated Masses on Mammograms of Dense Breasts. RSNA, Chicago, IL, November 27,1995.

9. Pisano ED, Hemminger BM, W. Garrett, E. Johnston, J. Chandromouli, D. Glueck, K. Muller, M. P. Braeuning, D. Puff, S. Pizer. Does CLAHE Image Processing Improve the Detection of Simulated Masses in Dense Breasts in a Laboratory Setting? Association of University Radiologists Meeting. Birmingham, AL. April 19,1996.

10. Pisano ED, Hemminger BM, W. Garrett, E. Johnston, S. Zong, D. Glueck, K. Muller, M. P. Braeuning, D. Puff, S. Pizer. Does CLAHE Image Processing Improve the Detection of Simulated Spiculations in Dense Breasts in a Laboratory Setting? Association of University Radiologists Meeting. Birmingham, AL. April 19,1996.

11. Pisano ED. The International Digital Mammography Development Group and the future of Digital Mammography. Meeting to Kick-off the National Library of Medicine Next Generation Internet Project to Develop a Digital Mammography Archive. University of Pennsylvania, Philadelphia, PA, November 12, 1998.

12. Pisano ED. Clinical Aspects of Digital Mammography. Update Course on Technical Aspects of Breast Imaging. Radiological Society of North America meeting. Chicago, IL. December 3, 1998.

13. Pisano ED. Current Status of Full Field Digital Mammography. Thomas Jefferson University Hospital. Department of Radiology. Philadelphia, PA. March 9, 1999.

14. Pisano ED. Current Status of Full Field Digital Mammography. American College of Radiology Imaging Network Semiannual Meeting. San Diego, CA. March 10, 1999.
15. Pisano E. Digital Mammography. Tenth Annual Excalibur Round Table Meeting for The American Cancer Society (national), Chapel Hill, NC, August 25,1995.
16. Hemminger B, Pisano ED, Johnston RE, et al. Mammographic Image Display Using a Workstation. Inforad Exhibitor (RSNA, 1996)
17. Hemminger B, Pisano ED, Johnston RE, et al. Workstation for Digital Mammography. Inforad Exhibitor (RSNA, 1997)
18. Hemminger BM, Pisano ED, Stahpit S, Johnston RE. Demonstration of Softcopy Display System for Digital Mammography. Radiologic Society of North America Meeting. Chicago, IL. November 29-December 4, 1998. Inforad Exhibitor (RSNA, 1998)

Patents and licenses applied for and/or issued

NA

Funding applied for based on work supported by this award:

1. Pisano E. Image Processing in Digital Mammography. RO1 renewal. Principal Investigator. To be Submitted to the National Cancer Institute Feb. 2000.
2. Pisano E. Image Processing for Digital Mammography. Principal Investigator. Awarded by The Susan Komen Foundation, beginning December 1, 1999.
3. Pisano, E. Tomosynthesis for Digital Mammography. Submitted to the Department of Defense, April 7, 1999.

Degrees obtained supported in part by this work:

1. Jayanthi Chandramouli. MS in Biomedical Engineering. UNC School of Medicine. Project title: The effect of Intensity Windowing on Detection of Simulated Breast Lesions in a Dense Breast in a Laboratory Setting. 1996.
2. Elodia B. Cole, MS in Biomedical Engineering, UNC School of Medicine. Project title: . Radiologist Preferences for Imaging Processing Algorithms for different clinical tasks for digital mammography display. (October 1999).
3. H. Zhong, PhD in Biomedical Engineering. UNC School of Medicine. Project title: Optimum Contrast Definition for Digital Mammography. In research phase.

Databases:

NA

Employment or research opportunities resulting from experience/training supported by this grant:

NA

Conclusions

The purpose of this research is to experimentally determine the diagnostic accuracy and clinical acceptability of digitally acquired mammograms displayed on soft copy display compared to laser printed hard copy. We have conducted observer studies both under laboratory conditions and under simulated conditions. We have used computer generated lesions and we have used real clinical mammograms to evaluate different image processing techniques. Our preliminary results indicate that digital images were preferred by radiologist observers to film screen radiographs for the diagnosis of masses with Unsharp masking processed mammograms statistically significantly preferred. For the screening task, film screen mammograms were preferred to all digital presentations, but Trex and MUSICA processed images were not statistically different in acceptability. For the calcification diagnostic task, no digital algorithm was preferred to film screen mammograms.

We are currently in the midst of conducting the observer experiment comparing diagnostic accuracy in a subset of the population, those patients presenting for problem-solving mammography, between soft copy display and images printed to film. Our preliminary observer study showed that the image processing tool was lesion type dependent. From the previous laboratory and clinical observers studies, we have been able to narrow the choice of image processing to manual intensity windowing for the film printed version and an automated histogram intensity windowing (two versions one that generates images that most closely resemble the film screen version and the other that best displays the dense breast areas, for the soft copy display.

Under sponsorship of this award and funding from other sources, we have acquired a library of about 380 digitally acquired mammograms.

We have developed a dual screen soft copy mammography workstation that is fast and user friendly. The observer study is presently underway and the results will be available by the end of the year.

References

There are no scientific references in the text of this report. Please refer to the attached papers in the appendices and the reference sections of each paper.

Personnel Receiving Payment

Faculty/Staff
Etta Pisano, MD
Eugene Johnston, PhD
Keith Muller, PhD
Bradley Hemminger, MS

Graduate Research Assistants:
Elodia Cole

Sanjay Sthapit
Shuquan Zong
Allan Dillon

Appendices

A.1

Pisano ED, Chandramouli J, Hemminger BM, DeLuca M, Glueck D, Johnston RE, Muller K, Braeuning MP, Pizer S. Does intensity windowing improve the detection of simulated calcifications in dense mammograms? Journal of Digital Imaging 1997;10(2):79-84.

A.2

Pisano ED, Chandramouli J, Hemminger BM, Johnston RE, Muller K, Pizer S. The effect of intensity windowing as an image processing tool in the detection of simulated masses embedded in digitized mammograms. Journal of Digital Imaging 1997;10(4):174-182.

A.3

Hemminger BM, Dillon A, Johnston RE, Muller K, Pisano ED, Deluca M. Evaluation of the effect of display luminance on the feature detection of simulated masses in mammograms. SPIE Medical Imaging 1997;3036:12.

A.4

Pisano ED, Zong S, Hemminger BM, DeLuca M, Johnston RE, Muller K, Braeuning MP, Pizer S. Contrast Limited Adaptive Histogram Equalization Image Processing to Improve the Detection of Simulated Spiculations in Dense Mammograms. Journal of Digital Imaging. 1998; 11(4): 193-200.

A.5

Pisano ED, Aylward S, Barbour P, Braeuning M.P, Brown ME, Chakraborty D, Cole E, Conant E, Eagle E, Fajardo LL, Feig S, Harrison J, Hemminger BM, R. Johnston RE, Jong R, Kennedy R, Kopans D, Kornguth P, Maidment A, Major S, McLelland R, Moore R, Muller K, Niklason L, Nishikawa R, Pizer SM, Plewes DB, Rosen E, Poyet C, Seaton K, Soo MS, Shumak R, Stahpit S, Staiger M, Vermont A, Walsh R, Williams MB, Williford M, Yaffe M, and Zong Z. Radiologist Preferences for Imaging Processing Algorithm for different clinical tasks for digital mammography display. Submitted to Radiology, Oct. 1999.

A.6

Pisano ED, Aylward S, Barbour P, Braeuning M.P, Brown ME, Chakraborty D, Cole E, Conant E, Eagle E, Fajardo LL, Feig S, Harrison J, Hemminger BM, R. Johnston RE, Jong R, Kennedy R, Kopans D, Kornguth P, Maidment A, Major S, McLelland R, Moore R, Muller K, Niklason L, Nishikawa R, Pizer SM, Plewes DB, Rosen E, Poyet C, Seaton K, Soo MS, Shumak R, Stahpit S, Staiger M, Vermont A, Walsh R, Williams MB, Williford M, Yaffe M, and Zong Z. Image Processing Algorithms for Digital Mammography – A Pictorial Essay. Submitted to Radiographics, Oct. 1999.

Appendix A.1

Pisano ED, Chandramouli J, Hemminger BM, DeLuca M, Glueck D, Johnston RE, Muller K, Braeuning MP, Pizer S. Does intensity windowing improve the detection of simulated calcifications in dense mammograms? Journal of Digital Imaging 1997;10(2):79-84.

The Effect of Intensity Windowing on the Detection of Simulated Masses Embedded in Dense Portions of Digitized Mammograms in a Laboratory Setting

Etta D. Pisano, Jayanthi Chandramouli, Bradley M. Hemminger, Deb Glueck, R. Eugene Johnston, Keith Muller, M. Patricia Braeuning, Derek Puff, William Garrett, and Stephen Pizer

The purpose of this study was to determine whether intensity windowing (IW) improves detection of simulated masses in dense mammograms. Simulated masses were embedded in dense mammograms digitized at 50 microns/pixel, 12 bits deep. Images were printed with no windowing applied and with nine window width and level combinations applied. A simulated mass was embedded in a realistic background of dense breast tissue, with the position of the mass (against the background) varied. The key variables involved in each trial included the position of the mass, the contrast levels and the IW setting applied to the image. Combining the 10 image processing conditions, 4 contrast levels, and 4 quadrant positions gave 160 combinations. The trials were constructed by pairing 160 combinations of key variables with 160 backgrounds. The entire experiment consisted of 800 trials. Twenty observers were asked to detect the quadrant of the image into which the mass was located. There was a statistically significant improvement in detection performance for masses when the window width was set at 1024 with a level of 3328. IW should be tested in the clinic to determine whether mass detection performance in real mammograms is improved.

KEY WORDS: mammography, breast mass, image processing

EFFECTIVE IMAGE display allows for an improvement in the clarity of structural details. Mammography, especially in patients with dense breasts, is a low-contrast examination that might benefit from increased contrast between malignant tissue and normal dense tissue. Image processing may allow for improved visualization of details within medical images.[1] Our overall aim is to improve the accuracy of mammography with image processing because 10% of palpable breast cancers are not visible with standard mammographic techniques.[2]

Contrast enhancement methods accentuate or emphasize particular objects or structures in an image by manipulating the gray levels in the display. This is done by imposing a predetermined transformation that amplifies the contrast between structures and effectively "resamples" the recorded intensities to enhance the properties of the displayed image.[3] These methods are not designed to increase or supplement the inherent structural information in the image, but simply improve the contrast and theoretically enhance particular characteristics.[4] Intensity windowing (IW) is an image processing technique that involves the determination of new pixel intensities by a linear transformation that maps a selected band of pixel values onto the available gray level range of the display system.[4]

Many investigators have studied the application of digital image processing techniques to mammography. McSweeney et al tried to enhance the visibility of calcifications by using edge detection for small objects, but never reported any clinical results.[5] Smathers et al showed that intensity band-filtering could increase the visibility of small objects compared to images without such filtering.[6] Chan et al used unsharp masking (an edge-sharpening technique used in photography for many years) to remove image noise for computerized detection of calcification clusters.[7] In another study, Chan et al noted that while these techniques improved detection, the improvements may have been greater if the observers had been trained to make diagnoses from the processed mammograms rather than the unprocessed (normal) mammograms.[8] Hale et al have applied nonspecific contrast and brightness adjustment through Adobe Photoshop (Adobe Systems Inc, Mountain View, CA) to digitized mammograms and have found improved performance by radiologists in determining the likelihood of malignancy of mammographically

From the Departments of Radiology, Computer Science, Biomedical Engineering, and Biostatistics, School of Medicine, School of Public Health, and College of Arts and Sciences, The University of North Carolina at Chapel Hill; and Picker International, St. David's, PA.

This research is supported by NIH PO1-CA 47982, NIH RO1-65583, and DOD DAMD17-94-J-4345.

Address reprint requests to Etta D. Pisano, MD, CB 7510, Room 503 Old Infirmary Building, Department of Radiology, UNC School of Medicine, Chapel Hill, NC 27599-7510.

0897-1889/97/1004-0006$5.00/0

apparent lesions.[9] Yin et al showed that nonlinear bilateral subtraction is useful in the computer-detection of mammographic masses.[10,11]

Previous work at the University of North Carolina has explored the use of intensity windowing (IW) and the Adaptive Histogram Equalization (AHE) family of algorithms in mammography and computed tomography.[12-14] We have previously described a laboratory-based method for testing the efficacy of an image processing algorithm in improving the detection of masses in dense mammographic backgrounds.[15] With that method, upon which our current work is based, radiologists and non-radiologists exhibit similar trends in detection performance. While non-radiologists did not perform as well as radiologists overall, the two populations displayed parallel increases and decreases in performance attributable to image processing.

The experiments described in this article were performed to determine whether IW could improve the detection of simulated masses in dense mammograms in a laboratory setting. Although the scope of this article is limited to the evaluation of observer performance using our established experimental paradigm, it may be interesting for follow-up work to evaluate these results with respect to measures proposed by other authors, such as the conspicuity measure proposed by Revesz et al and Revesz and Kundel.[16-18]

MATERIALS AND METHODS

The experimental paradigm reported here is based on the model we have previously described and allows for the laboratory testing of a range of parameter values (in this case, window width and level).[15] The experimental subject is shown a series of test images that consist of an area of a dense mammogram with a simulated mass embedded in the image in one of its four quadrants. The observer's task is to determine in which quadrant the mass is located. The test images are displayed in both the processed and unprocessed format, and the contrast of the object is varied, from quite easy to detect to impossible to detect.

A computer program randomly selected one of 40 background images and rotated that background to one of four orientations. The 40 background images of 256 × 256 pixels each were extracted from actual clinical film screen mammograms digitized using a Lumisys digitizer (Lumisys Inc, Sunnyvale, CA) with a 50 micron sample size with 12 bits (4096 values) of density data per sample. The images contained relatively dense breast parenchyma. These were determined to be dense by a radiologist expert in breast imaging. Only areas that contained relatively uniformly dense tissue were included, with adjacent fatty areas specifically excluded. These areas were selected because they are most likely to hide soft tissue masses in the clinical setting. They were known to be normal by virtue of at least three years of normal clinical and mammographic follow-up. They were selected by a breast imaging radiologist from digitized film screen craniocaudal or mediolateral oblique mammograms. Figure 1 shows one of the backgrounds. The density of this background as displayed in this figure is typical of those used in the experiments.

These 40 images and four orientations provided 160 different dense backgrounds. Next, the program added a phantom feature (a mass) into the background. The image was processed with IW to yield the final stimulus.

Mammographic masses were simulated by blurring (through convolution with a gaussian kernel with a standard deviation of 2.0 pixels) a disk that is approximately 5 mm in diameter when printed on film (1.51 degree visual angle at a 38 cm viewing distance). The masses were added at four fixed contrasts. The four contrasts added were, in digitized density units, 20, 40, 80, and 160 digital driving levels (DDLs). Although contrast is commonly defined as a change in luminance with respect to the background luminance, we used only the change in luminance in this experiment because the change was independent of the background luminance. This is because contrast was represented in log luminance (ie, the DDLs corresponded to optical density), and since all the study backgrounds were in the luminance range where Weber's law holds, adding a mass of constant density equates to a constant change in contrast, independent of the background luminance. DDL's do *not* correspond directly to just noticeable differences (JNDs). In fact, they correspond to fractions of JNDs for the case of the display system used in these experiments.

Although the simulated structures were not entirely realistic, they did, however, possess the same scale and spatial characteristics of actual masses typically found at mammography. Figure 2 shows an example of a simulated mass. Figure 3 shows a typical background image with the mass added to it. We used simulated features instead of real features so that we could have precise control over the location, orientation, and figure-to-background contrast of the masses.

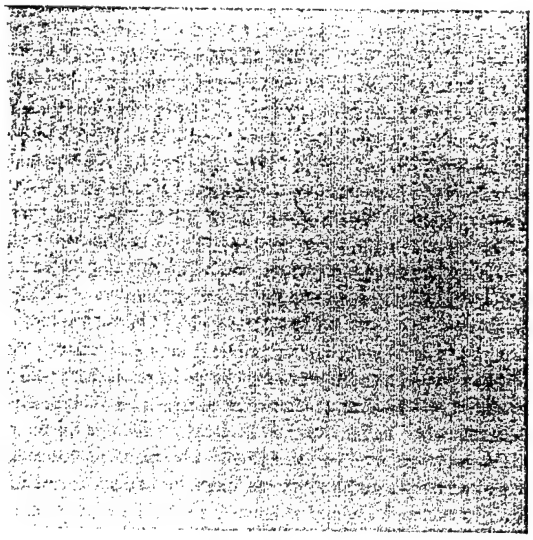

Fig 1. An example of a dense normal background taken from a patient's mammogram and used in the reported experiments.

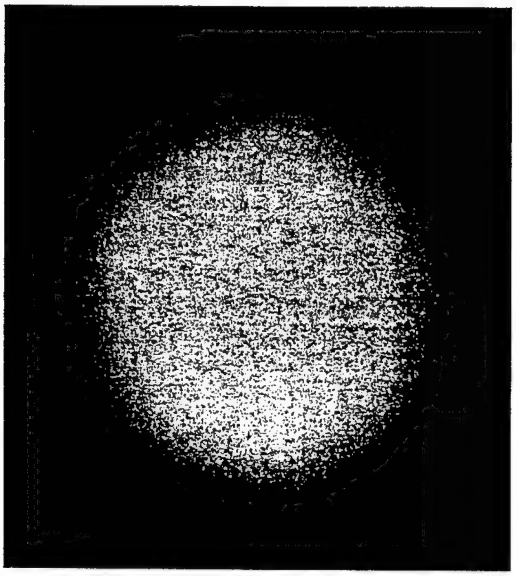

Fig 2. An example of a simulated mass. The actual size of the masses used in the experiments was only 5 mm.

A 3 × 3 grid of window and level parameters was designed based on the results of pilot preference studies done with two radiologists who specialize in breast imaging. In these pilot studies, the two radiologists reviewed dense mammograms with real clinical lesions that were judged to be difficult to visualize using standard film screen mammography. There were 7 images of this type reviewed with 70 combinations of window width and level applied. The radiologists scored each combination of values as showing no change over the standard image, improving the visibility of the lesion, or worsening its visibility.

For experiment 1, the grid spanned all the likely optimal settings (windows of 512, 768, 1024 and levels of 3072, 3328, 3584). Thus, there were a total of 10 IW settings (including the default unprocessed image, with a window width of 4096 and level of 2048) that were applied throughout experiment 1.

To confirm the results of the first experiment and to examine additional IW settings, experiment 2 was performed. Experiment 2 also included the unprocessed (wide open window width) condition and 9 other IW conditions. The combinations of parameters evaluated in Experiment 2 were as follows: window width of 640 with levels of 3456, 3584 and 3840; window width of 1024 with levels of 3200, 3328 and 3584; and window width of 1536 with levels of 2944, 3072, and 3328).

The digital images were printed onto standard 14 × 17 inch single emulsion film (3M HNC Laser Film; 3M, St. Paul, MN) using a Lumisys Lumicam film printer (Lumisys). Each original 50 micron pixel was printed at a spot size of 160 microns, which produced 4 × 4 centimeter film images, resulting in an enlargement by a factor of 3.2. The background and target are magnified together. The radiologist observers in the pilot experiment reported that the magnification did not make the backgrounds unrealistic. Forty images were printed per sheet of film. The images were randomly ordered into an 8 × 5 grid on each sheet of film. Both the film digitizer and film printer were calibrated, and measurements of the relationship between optical density on film and digital units on the computer were determined to generate transfer functions describing the digitizer and film printer. To maintain a linear relationship between the optical densities on the original analog film and the digitally printed film, we calculated a standardization function that provided a linear matching between the digital and printer transfer functions. This standardization function was applied when printing the films to maintain consistency between the original optical densities of the original mammography film and those reproduced on the digitally printed films. The film printer produces films with a constant relationship between an optical density (OD) range of 3.35 to 0.13, corresponding to a digital input range of 0 to 4095, respectively.

There were 20 observers for each experiment. These were graduate students from the medical school, biomedical engineer-

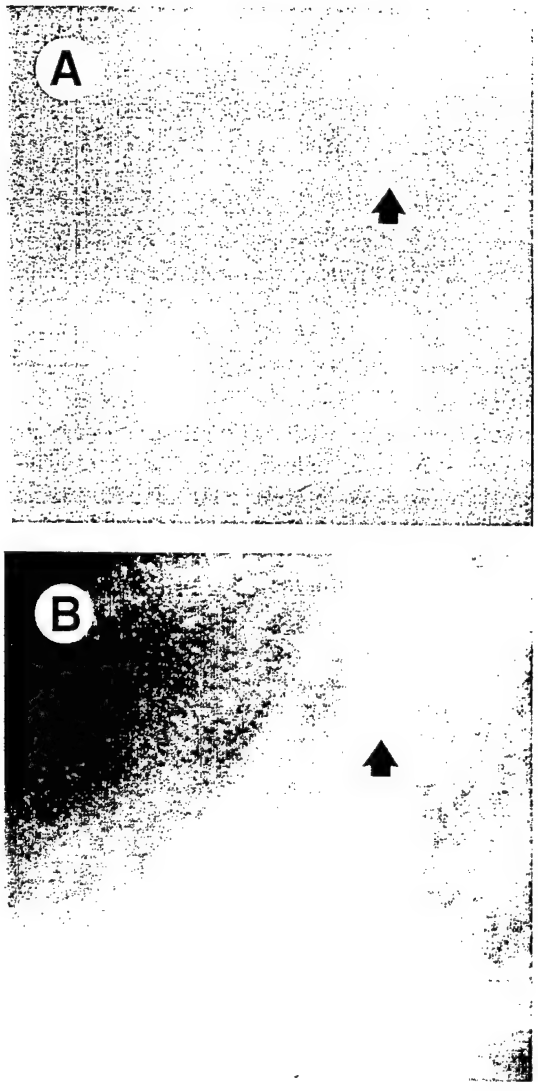

Fig 3. A dense background with a simulated mass embedded in it in the right upper quadrant in both figures (arrow). (A) is the default unprocessed image with window width 4096 and level 2048. (B) is the same image with window width 1024 and level 3328.

ing department, and computer science department. Performance bonus pay was provided. Observers selected the quadrant of the image that they thought contained the mass. All images contained a mass. Observers were told to make their best guess if they could not see the simulated mass with certainty.

Films were displayed in a darkened room on a standard mammography lightbox that was masked so that only the grid of images on the film was illuminated. Observers could move closer to the image and could use a standard mammography magnifying glass, as desired. The observers were trained for the task through the use of two sets of stimulus image films with instructive feedback before actually starting the experiment.

Both experiments had the same basic design. The order of the presentation of the stimuli was counterbalanced so as to eliminate any systematic effect of unimportant variables. All 160 possible combinations of processing condition (10 IW levels), contrast level (4 contrasts) and location of the masses (4 quadrants) were used in the experiment. The experiment was designed to have 5 self-contained blocks, in which all 160 combinations appeared. The intent was to have the observer see all the combinations in each block in case the observer was unable to complete the experiment. In fact, all observers did complete the experiment. There were 40 backgrounds and 4 possible rotations of each background, for 160 possible background patterns. For each block, a different background pattern was assigned uniquely to each of the 160 possible combinations. The assignment was different for each block. Each observer looked at a total of 800 images, which were the 160 possible combinations, each superimposed on 5 backgrounds.

Observers were instructed to take breaks after each block of stimuli, and more often if necessary. No time limit was imposed on the observers viewing duration of the test images. Overall, the experiment took 2 hours for each observer, divided into two sessions of approximately 60 minutes each. The two sessions were always scheduled on two different days within a week of each other.

Data Analysis Overview

Classical sensory discrimination theory predicts that because contrast values were varied from virtually imperceptible to highly apparent, a typical S-shaped curve will describe the data.[2] At values where the contrast was very low, on average observers will guess randomly and get approximately 25% right because there are four choices. Where the contrast is very high, they will almost always get the correct answer. This relationship between log10 of the contrast of the object relative to the background intensity and the percent correct can be described with a probit model. This model is typically used to describe the relationship between a continuous predictor (log contrast) and a discrete variable (percent correct), and assumes that the curve between them is described by the cumulative gaussian distribution.

Probit models were fit for each subject and enhancement condition using contrast (DDLs of mass above background) as the predictor. The probability that a subject gets a correct answer is given by the following equation:

$$\Pr[\text{correct}] = \tfrac{1}{4} + (1 - \tfrac{1}{4})\,\Phi\,[(x - \mu_{ij})/\sigma_i$$

Here i indexes subjects, and j indexes enhancements with x representing the log (contrast). Classical psychophysical theory and experimental results strongly support the use of the logarithmic transform, as did our data. In the experiments reported here, we used x = log10 (number of DDLs above background). The subscripts in the equation indicate that for each subject a single spread parameter was estimated (which pools across all stimuli and conditions). Also, for each subject, a separate location parameter was estimated for each enhancement condition. With 10 processing conditions, this implies a total of 10 location parameter estimates and one spread parameter for each subject. Our assumption, that there is a common spread parameter, makes sense biologically because it corresponds to linearity of the perceptual mapping. It is advantageous to an organism to have the same amount of change in stimulus produce a constant perceptual response, and that is precisely how the human visual system works over a wide range.

The location parameter (μ) is the mean of the corresponding gaussian distribution and the inflection point of the sigmoidal probit curve. Processing conditions that improve detection will cause this parameter to be smaller, and the curve will shift to the left, or equivalently if viewed from the perspective of the same contrast value, the curve shifts upward. This occurs because lower contrast levels are required to spot the object. When the processing of the image makes detection harder, higher contrast levels are needed to locate the mass, and the curve shifts to the right. The values of σ, the spread parameter, correspond to the slope of the line. Large values of σ correspond to steep slopes.

The probit analysis summarized the relationship between contrast and proportion correct for each subject and processing condition. To compare the processing conditions and to examine the effect of window width and level, further analysis was needed. To include both the mean and the location parameter from the probit analysis, we defined an overall measure to be $\theta_{ij} = \mu_{ij} + \sigma_i$, which corresponds to 88% correct. Because we were interested in the improvement offered by IW, we measured the "success" of a processing condition by calculating the difference between its θ score and the θ score for the unprocessed image for each subject. A large positive difference of θ score reflects improved performance because it indicates better detection with processed images than with unprocessed images.

For each experiment, two analyses were performed using this outcome measure. To keep an overall experiment-wide type 1 error rate of .05, a repeated measures analysis of variance (ANOVA) was done at the .04 level, with a set of nine *t*-tests at the .01/9 level.

Repeated measures analysis of variance is a technique used to analyze data in which many measurements were made on each subject. It allows one to examine the effect of processing conditions and their interactions, while allowing for the dependence of measurements taken on the same observers. With the difference in θ scores as the outcome, and window width and level as the predictors, the repeated measures ANOVA model was fitted.

The model can be thought of as a response surface in three dimensions with performance plotted against window width and level. A flat surface would mean that window width and level had no effect on the outcome. The major hypothesis tested in the ANOVA is equivalent to asking the question, "Is the response surface flat?" If it is not flat, the step-down hypotheses allow one to ask what shape the surface is, whether it is curved in both directions (quadratic by quadratic trends), curved in one direction and sloped in the other (quadratic by linear trends), or sloped in both directions (linear by linear trends). A peak in the

surface means that there is one image processing technique that is better than any other. Conversely, if the difference score is equal to zero for any intensity windowing setting, it would correspond to no difference between the processed image and the unprocessed image. That is what the *t* statistics test.

RESULTS

Experiment 1

The repeated measures ANOVA revealed that there was a significant interaction between window width and level ($P = .0001$, G-Gê = .8347). To examine the nature of this interaction, a series of step-down tests was planned. There was a significant interaction between a quadratic trend in window width and a quadratic trend in level ($F = 31.08$, $P = .0001$). Because the quadratic by quadratic interaction was significant, no further tests were examined. A quadratic by quadratic trend means that the surface was curved with respect to both window width and level, and that the shape of the curve differed for fixed levels of window width and level (Figs 4 and 5).

At the overall .01 level, the differences between the enhancement conditions and the unenhanced were examined. The null hypothesis is that there will be no difference between the mean θ for the unenhanced and an enhancement condition. There are nine such hypotheses, corresponding to the nine enhancements. A Bonferroni correction to control the overall error rate made each individual α level .0011. Four settings of IW made finding the masses significantly harder, three made the task significantly easier, and two made no significant difference. The settings that made the task easier are window width 1024 with level 3328, window width 768 with level 3584 and window width 1024 with level of 3584 (Table 1).

Experiment 2

Again, the repeated measures ANOVA showed that there was significant interaction between window width and level ($P < .0001$, $F = 60.9$; Figs 6 and 7). As in experiment 1, a quadratic by quadratic interaction was significant ($P < .0001$, $F = 32.61$). Table 2 shows the results of nine two-sided *t*-tests. Only one image processing setting resulted in significantly better performance than the unprocessed, namely window width of 1024 with a window level of 3328 ($P < .0001$). Seven of the settings were not significantly different from the unprocessed image. One setting was significantly worse (Table 2).

The probit model predicts that IW will increase detection of masses. For example, at the contrast level of 40 DDLs above background, which is the contrast level tested that was nearest to the observer's detection threshold, these results predict that the feature detection rate would change from 51% to 68% for the conditions of experiment 1, and from 52% to 67% for the conditions of experiment 2 (Figs 5 and 7).

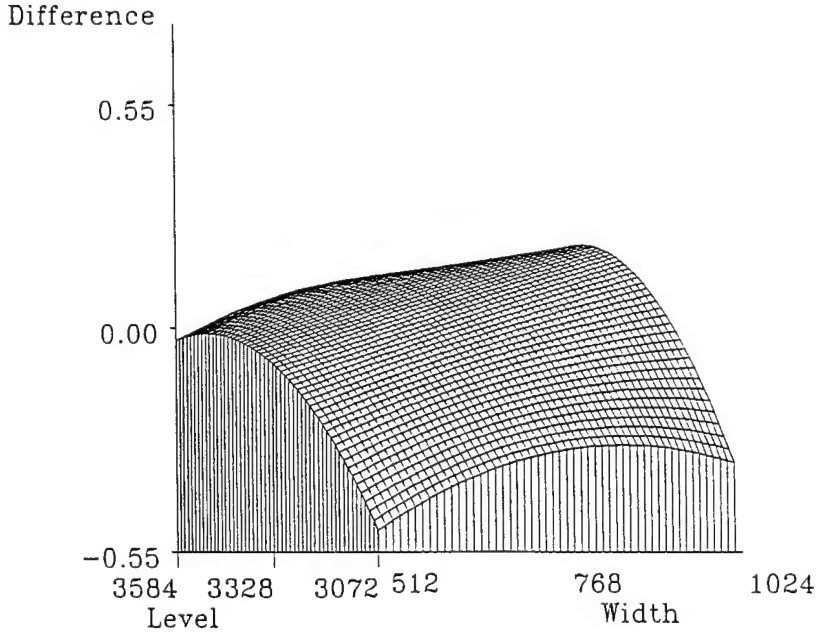

Fig 4. Interpolated predicted values from repeated measures ANOVA for Study 1, the difference in θ value versus window width and window level.

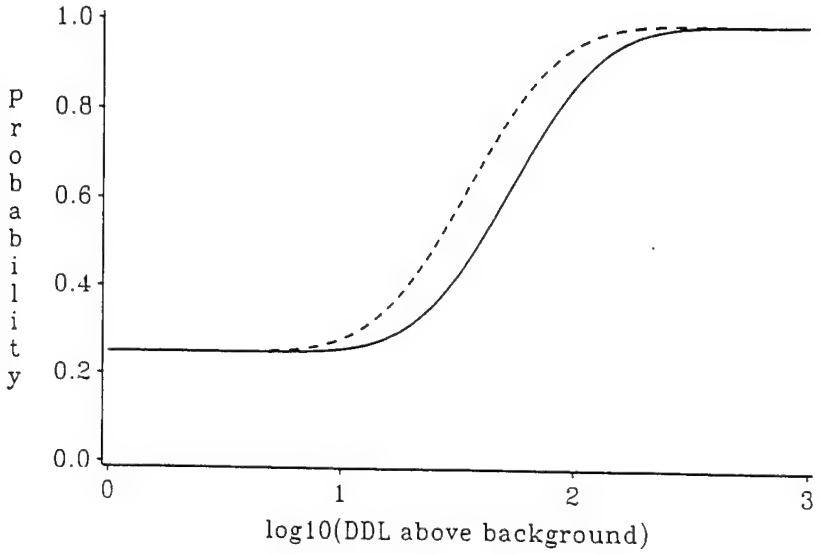

Fig 5. Estimated detection probability from Study 1 for window width of 1024 and window level of 3328 (---) versus unprocessed (—) condition. The shift in the curve to the left reflects improved detection.

DISCUSSION

These results are encouraging. This is the first experiment in mammography that demonstrates that an algorithm can improve the detection of a simulated mass placed in a dense mammogram. At the same time, it is obviously important to choose the window width and level with care because performance can be significantly degraded if inappropriate parameters are chosen.

What do these results mean for clinical mammographers? Will we be using this technology in the clinic in detecting lesions in dense mammograms? The use of graduate student observers and the use of simulated masses in this study might incorrectly predict the performance of radiologists in detecting real masses in real patients. We have demonstrated previously that graduate student performance at this task parallels the performance of experienced mammographers.[15] Evaluation by radiologists on real patients will determine the ultimate utility of this algorithm in the clinical setting. Because we have used real clinical images and we have simulated masses using relatively realistic stimuli, we are optimistic that these methods will improve clinical performance and that radiologists will be using IW to help them in determining whether mammograms of women with dense breasts really do contain masses.

Table 1. Summary of differences between unenhanced and enhanced θ for Study 1

Window Level	Window Width	Mean Difference in θ	*SD*	*P* value
3072	512	−.50	.108	.0001
3072	768	−.32	.093	.0001
3072	1024	−.34	.089	.0001
3328	512	−.11	.074	.0001
3328	768	.04	.087	.0706
3328	1024	.18	.104	.0001
3584	512	−.03	.097	.1716
3584	768	.14	.082	.0001
3584	1024	.12	.121	.0004

Note: Positive values in mean difference in θ column correspond to improved detection of simulated masses.

One could argue that our methods are limited because the small areas studied make IW more useful than it would be in larger areas. By magnifying the original 12.8 mm × 12.8 mm image to 40 mm × 40 mm during the printing process, the variation in density may be reduced compared to the variation of an actual 40 mm × 40 mm cropped section of a mammogram, because a third fewer samples are included. In a similar experiment,[19] we found that the variation difference between cropped mammographic sections of different sizes from uniformly dense areas of mammograms was small, and unlikely to have a significant effect on feature detection of masses when using this experimental paradigm. In addition, ideally one would report on the standard deviations of the σ of the pixel values of the background as a parameter affecting the probability of detection of the mass embedded in the background. Although we report this data in all other experiments using this paradigm, unfortunately, we are unable to do so for this experiment owing to an error by the programmer.

Digital mammography will be available in the clinic very soon. It is obvious that image processing

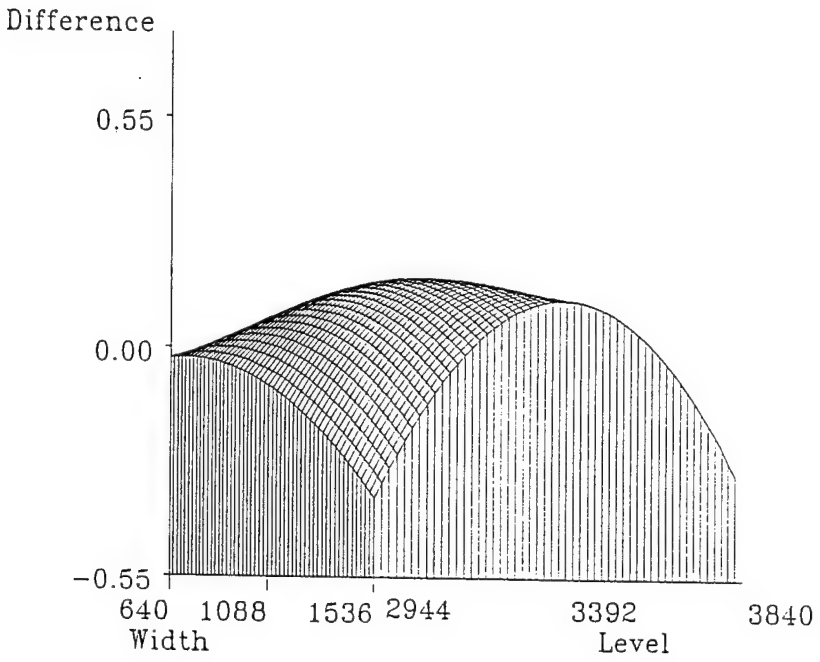

Fig 6. Interpolated predicted values from repeated measures ANOVA for Study 2, the difference in θ value versus window width and window level.

will be used to optimize the visibility of lesions in digital mammograms.[20] Ideally, any image processing algorithm that might be useful will be tested on real patients in that setting. That will be an expensive and time consuming process that will involve real patients making clinically important decisions about their own breast health, including the advisability of biopsy, lumpectomy, and mastectomy. Ideally, before this technology arrives in the clinic, radiologists will have some idea of which category of algorithms to test in that setting. This work is intended to give radiologists preliminary data to narrow the choices that might be useful before the expensive clinical tests are undertaken. This approach suggests not only which algorithms might help clinically but which parameter settings most improve detection.

One could take the approach that the IW dials should be spun until a clinically pleasing image is displayed. This approach might be acceptable and even convincing to many radiologists. It is at least possible that what pleases radiologists in terms of the aesthetics of the image might not improve the detection performance of their visual systems, and

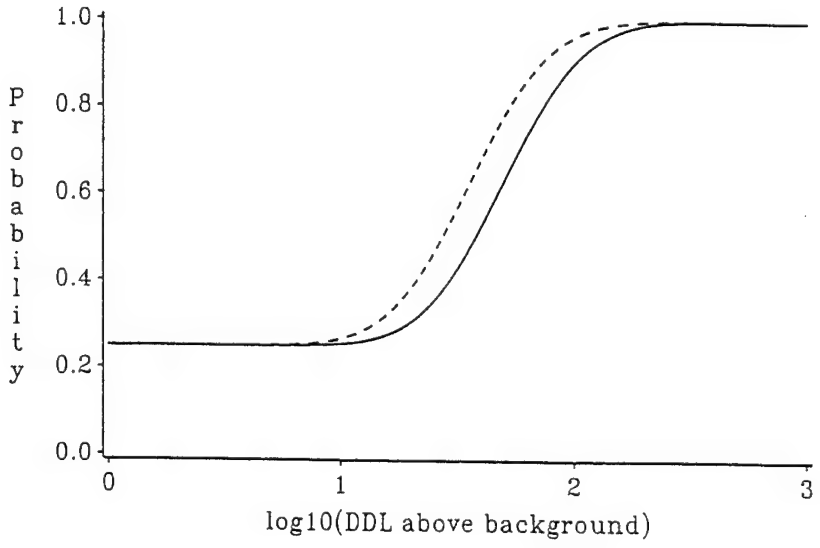

Fig 7. Estimated detection probability from Study 2 for window width of 1024 and window level of 3328 (---) versus unprocessed (—) condition. The shift in the curve to the left reflects improved detection.

Table 2. Summary of differences between unenhanced and enhanced θ for Study 2

Window Level	Window Width	Mean Difference in θ	*SD*	*P* value
3456	640	0.04	0.08	.0239
3584	640	−0.05	0.09	.0215
3840	640	−0.31	0.09	.0001
3200	1024	0.04	0.07	.0142
3328	1024	0.14	0.08	.0001
3584	1024	0.01	0.09	.6155
2944	1536	−0.02	0.07	.1255
3072	1536	0.06	0.08	.0045
3328	1536	0.06	0.07	.0013

Note: Positive values in mean difference in θ column correspond to improved detection of simulated masses.

in fact, could worsen their detection performance. This project was intended to be more rigorous in exploring the window widths and levels that might be useful in the most challenging areas of the breast, namely the dense parts. We also have performed similar experiments on the AHE class of algorithms.[21,22]

This experiment does not address how IW would effect the appearance of fatty areas of the breast, and the delectability of lesions in those parts. We would not want to apply an algorithm that degrades performance in areas of the breast where sensitivity is quite high with current technology. There are two possible technical responses to that concern. First, IW could be applied selectively to only the dense areas as an adjunct to the more standard appearing mammogram with the radiologist pointing and clicking to the areas where windowing would be desirable. Alternatively, the IW could be individualized to the patient's unique intensity histogram so that the areas to be processed of the image could be selected by the computer itself. In fact, ideally the computer could be programmed to choose an individual IW setting for each portion of the mammogram so that contrast was preserved in all portions of the image. Ongoing experiments in our laboratory are currently exploring the latter possibility.

Of course, our results to date cannot estimate the exact frequency of false-positive diagnoses when IW is used. Many alternate forced choice tests (in our case, 4-AFC) yield proportion correct as the primary outcome. MacMillan and Creelman discussed methods for converting proportion correct in this setting to a value of d′, the sensitivity parameter of an receiver operating characteristic (ROC) analysis.[23] The particular choice of conversion depends on side conditions concerning the nature of any rater basis. Given the characteristics of the study design, subjects and training, we believe that superior proportion correct will translate into superior d′. If this is true, the practical value of IW must be tested in a clinical setting. Then ROC analysis will allow separate analysis of a reader's sensitivity and pay off function on the performance of the technique as part of a diagnostic system.

CONCLUSION

The testing of these methods on patients with palpable and mammographically detected lesions has been funded by the National Cancer Institute and the Department of Defense, and will be ongoing over the next few years at the University of North Carolina and Thomas Jefferson University Hospital. We expect to evaluate both IW and Contrast Limited Adaptive Histogram Equalization (CLAHE) in the clinical setting to determine whether or not these algorithms improve the performance of radiologists in detecting and characterizing breast lesions.

REFERENCES

1. Rosenman J, Roe CA, Cromartie R, et al: Portal film enhancement: Technique and clinical utility. Int J Radiat Oncol Biol Physics 25:333-338, 1993

2. Homer MJ: Mammographic interpretation: A practical approach. New York, NY, McGraw Hill, 1991, pp 4-5

3. Pizer SM: Psychovisual issues in the display of medical images, in Hoehne KH (ed): Pictorial Information Systems In Medicine. Berlin, Springer-Verlag, 1985, pp 211-234

4. Jain AK: Fundamentals of Digital Image Processing. Englewood Cliffs, NJ, Prentice Hall, 1989

5. McSweeney MB, Sprawls P, Egan RL: Enhanced image mammography. AJR Am J Roentgenol 140:9-14, 1983

6. Smathers RL, Bush E, Drace J, et al: Mammographic microcalcifications: Detection with xerography, screen-film, and digitized film display. Radiology 159:673-677, 1986

7. Chan HP, Doi K, Galhorta S, et al: Image feature analysis and computer-aided diagnosis in digital radiography: 1. Automated detection of microcalcifications in mammography. Med Phys 14:538-547, 1987

8. Chan HP, Vyborny CJ, MacMahon H, et al: Digital mammography ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. Invest Radiol 22:581-589, 1987

9. Hale DA, Cook JF, Baniqued Z, et al: Selective Digital Enhancement of Conventional Film Mammography. J Surg Oncol 55:42-46, 1994

10. Yin F, Giger ML, Vyborny CJ, et al: Comparison of Bilateral-Subtraction and Single-Image Processing Techniques in the Computerized Detection of Mammographic Masses. Invest Radiol 28:473-481, 1993

11. Yin F, Giger M, Doi K, et al: Computerized detection of masses in digital mammograms: Analysis of Bilateral Subtraction Images. Med Phys 18:955-963, 1991

12. Hemminger BM, Johnston RE, Muller KE, et al: Comparison of clinical findings between intensity-windowed versus CLAHE presentation of chest CT images. SPIE Medical Imaging VI: Image Capture, Format and Display 1653:164-175, 1992

13. Pizer SM, Zimmerman JB, Staab EV: Adaptive grey level assignment in CT scan display. J Comput Assist Tomog 8:300-305, 1984

14. Puff DT, Cromartie R, Pisano ED: Evaluation and optimization of contrast enhancement methods for medical images. Proceedings of the SPIE Visualization in Biomedical Computing Conference 1808:336-346, 1992

15. Puff DT, Pisano ED, Muller KE, et al: A method for determination of optimal image enhancement for the detection of mammographic abnormalities. J Digit Imaging 7:161-171, 1994

16. Revesz G, Kundel HL, Graber MD: The influence of structured noise on the detection of radiologic abnormalities. Invest Radiol 9:479-486, 1974

17. Kundel HL, Revesz G: Lesion conspicuity, structured noise and reader error. AJR Am J Roentgenol 126:1233-1238, 1976

18. Revesz G, Kundel HL: Psychophysical studies of detection errors in chest radiology. Radiology 128:559-562, 1977

19. Hemminger BM, Johnston RE, Rolland JR, et al: Introduction to perceptual linearization for video display systems for medical image presentation. J Digit Imaging 8:21-34, 1995

20. Shtern F: Digital mammography and related technologies: A perspective from the National Cancer Institute. Radiology 183:629-30, 1992

21. Pisano ED, Hemminger BM, Garrett W, et al: Does CLAHE image processing improve the detection of simulated masses in dense breasts in a laboratory setting? Presented at the Association of University Radiologists Meeting. Birmingham, AL, April 19, 1996

22. Pisano ED, Hemminger BM, Garrett W, et al: Does CLAHE image processing improve the detection of simulated spiculations in dense breasts in a laboratory setting? Presented at the Association of University Radiologists Meeting. Birmingham, AL, April 19, 1996

23. MacMillan NA, Creelman CD: Detection theory: A user guide. Cambridge, UK, Cambridge University Press, 1991, pp 135-136

Appendix A.2

Pisano ED, Chandramouli J, Hemminger BM, Johnston RE, Muller K, Pizer S. The effect of intensity windowing as an image processing tool in the detection of simulated masses embedded in digitized mammograms. Journal of Digital Imaging 1997;10(4):174-182.

Does Intensity Windowing Improve the Detection of Simulated Calcifications in Dense Mammograms?

Etta D. Pisano, Jayanthi Chandramouli, Bradley M. Hemminger, Marla DeLuca, Deb Glueck, R. Eugene Johnston, Keith Muller, M. Patricia Braeuning, and Stephen Pizer

This study attempts to determine whether intensity windowing (IW) improves detection of simulated calcifications in dense mammograms. Clusters of five simulated calcifications were embedded in dense mammograms digitized at 50-μm pixels, 12 bits deep. Film images with no windowing applied were compared with film images with nine different window widths and levels applied. A simulated cluster was embedded in a realistic background of dense breast tissue, with the position of the cluster varied. The key variables involved in each trial included the position of the cluster, contrast level of the cluster, and the IW settings applied to the image. Combining the ten IW conditions, four contrast levels and four quadrant positions gave 160 combinations. The trials were constructed by pairing 160 combinations of key variables with 160 backgrounds. The entire experiment consisted of 800 trials. Twenty student observers were asked to detect the quadrant of the image in which the mass was located. There was a statistically significant improvement in detection performance for clusters of calcifications when the window width was set at 1024 with a level of 3328, and when the window width was set at 1024 with a level of 3456. The selected IW settings should be tested in the clinic with digital mammograms to determine whether calcification detection performance can be improved.

KEY WORDS: mammography, image processing, intensity windowing, observer studies, calcifications, computers, radiology.

MAMMOGRAPHY, especially in women with dense breasts, is not perfectly sensitive to all cancers. Approximately 10% to 15% of palpable malignancies are not visible mammographically.[1] There is some reason to believe that digital mammography might allow for greater contrast and improved detection of small and early tumors over standard film screen technology, especially if image processing is used to improve image contrast.[2,3]

There are many potentially useful image processing algorithms, and each algorithm has a number of parameters that can be systematically varied to improve or worsen lesion detectability. Radiologists cannot and should not evaluate these algorithms in the clinic with real patients. Such a task would be overwhelming and potentially could cause much unnecessary patient anxiety. Ideally, a test set of image phantoms with simulated lesions in known locations should be used to test each potentially useful algorithm and its attendant parameters in the laboratory setting before any patient's images are interpreted using these algorithms. We have developed such a laboratory method for evaluation of image processing algorithms.[4] In previous work, we have shown that detection performance with the application of contrast limited adaptive equalization (CLAHE) to digitized mammograms is parallel for radiologists and student observers.[4] Using the same experimental paradigm, we report here on whether intensity windowing (IW) can improve the detection of calcifications in dense mammograms in a laboratory setting. We have previously reported elsewhere that IW improves the detectability of masses in dense mammograms.[5]

Many investigators have studied the use of image processing techniques in digitized mammograms. McSweeney attempted to improve the visibility of calcifications by using edge detection for small objects, but gave no clinical results.[6] Smathers improved the visibility of small objects in images by intensity band-filtering.[7] Chan used unsharp-masking to reduce image noise to improve detection of clustered calcifications.[8] Chan, Hale, and Yin have tested other image processing methods on digitized mammograms with variable results.[9-12]

Contrast enhancement methods are not designed to increase or supplement the inherent structural information in an image, but rather to improve the image contrast and theoretically to enhance particular characteristics. IW is an image processing technique that involves the determination of new

From the Departments of Radiology, Computer Science, Biomedical Engineering, and Biostatistics, The University of North Carolina-Chapel Hill, School of Medicine, School of Public Health and College of Arts and Sciences, Chapel Hill, NC.

Supported by NIH PO1-CA 47982, NIH RO1-65583 and DOD DAMD 17-94-J-4345.

Address reprint requests to Etta D. Pisano, MD, CB 7510, Room 503 Old Infirmary Building, Dept of Radiology, UNC School of Medicine, Chapel Hill, NC 27599-7510.

0897-1889/97/1002-0002$3.00/0

pixel intensities by a linear transformation that maps a selected band of pixel values onto the available gray level range of the display device.[13]

The experiments described in this article were performed to determine whether IW could improve the detection of simulated clusters of calcifications in dense mammograms in a laboratory setting. Although the scope of this article is limited to the evaluation of observer performance with respect to the contrast of the simulated microcalcification to background using our established experimental paradigm, it may be interesting for follow-up work to evaluate these results with respect to measures proposed by other investigators, such as the conspicuity measure proposed by Revesz and Kundel.[14-16]

MATERIALS AND METHODS

The experimental paradigm used here is based on the model we have previously described and allows for the laboratory testing of a range of parameter values (in this case, window width and level).[4] The experimental subject is shown a series of test images that consist of an area of a dense mammogram with a simulated cluster of calcifications embedded in the image in one of four quadrants. The observer's task is to determine in which quadrant the cluster of calcifications is located. The test images are displayed in both the processed and unprocessed format, and the contrast of the object against the background is varied from quite easy to detect to impossible to detect.

A computer program randomly selected one of 40 background images and rotated that background to one of four orientations. The 40 backgrounds images of 256 × 256 pixels each were taken from actual mammograms that had been digitized using a Lumiscan digitizer (Lumisys, Inc, Sunnyvale, CA) with a 50 μm sample size and 12 bits of intensity data per sample. The images were selected from relatively dense parts of the mammograms that were known to be normal by virtue of 3 years of clinical and mammographic follow-up. They were selected by a radiologist expert in breast imaging from digitized film screen craniocaudal or mediolateral oblique mammograms. Fig 1 shows one of the backgrounds.

The gray scale values for the mammographic backgrounds are assigned the values recorded by the Lumisys digitizer. The digitizer assigns digital values in the range 495 to 4095 representing an optical density range of 3.43 to 0.08. The digitizer produces digitized gray values that map one to one with optical density (OD) values, ie, the same OD value on film will produce the same gray level.

The 40 images and four orientations provided 160 different dense backgrounds. The program then added a phantom feature, the simulated cluster of five calcifications into the background. The image was then processed with IW to yield the test stimulus.

Mammographic calcifications were simulated using a locally developed program. A cluster of five calcifications was generated. Each individual calcification was a square measuring 1 pixel by 1 pixel in size. Simulated clusters were used instead of real features so that we could have precise control over the structure location, orientation, and structure to background contrast of the calcifications. To more realistically simulate microcalcifications would have required using multiple pixels per microcalcification, for instance a 2 × 2 or 3 × 3 matrix. Because the smallest spot size available to use at the time for printing films was 160 μm per pixel, the use of a 2 × 2 or 3 × 3 microcalcification would have unrealistically enlarged the simulated microcalcification. Thus we limited our simulated calcifications to single pixel areas, and varied only the contrast of the calcification. As a result, the simulated calcifications were not entirely realistic, but they did possess the same scale and similar spatial characteristics to actual calcifications seen at mammography.

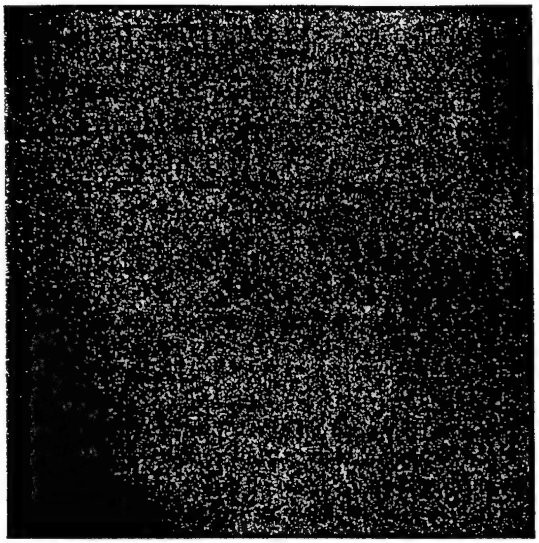

Fig 1. An example of a dense normal background taken from a patient's mammogram and used in the experiment.

The intensity difference of the calcifications from background was defined as the gray level of the digital microcalcifications before addition to the background. The calcifications were then embedded at four different intensity levels equally spaced in perceived brightness relative to background by pixel-wise addition of the structure and background images. Fig 2 shows an example of a simulated cluster of calcifications. Figure 3A shows a typical background with the cluster embedded in it without windowing applied. Figure 3B shows the same image with intensity windowing, with the window width of 1024 and a level of 3328. The images in Figs 2 through 3 were photographed from a video monitor with a larger pixel spot size.

A 3 × 3 grid of appropriate window and level parameter settings was selected based on the results of pilot preference

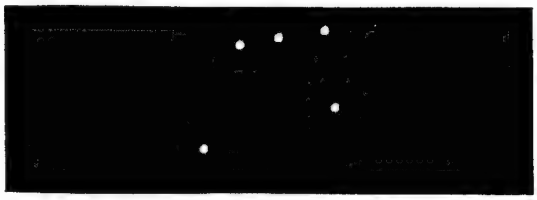

Fig 2. An example of a simulated cluster of calcifications. The actual size of the cluster used in the experiment was only 5 mm in diameter.

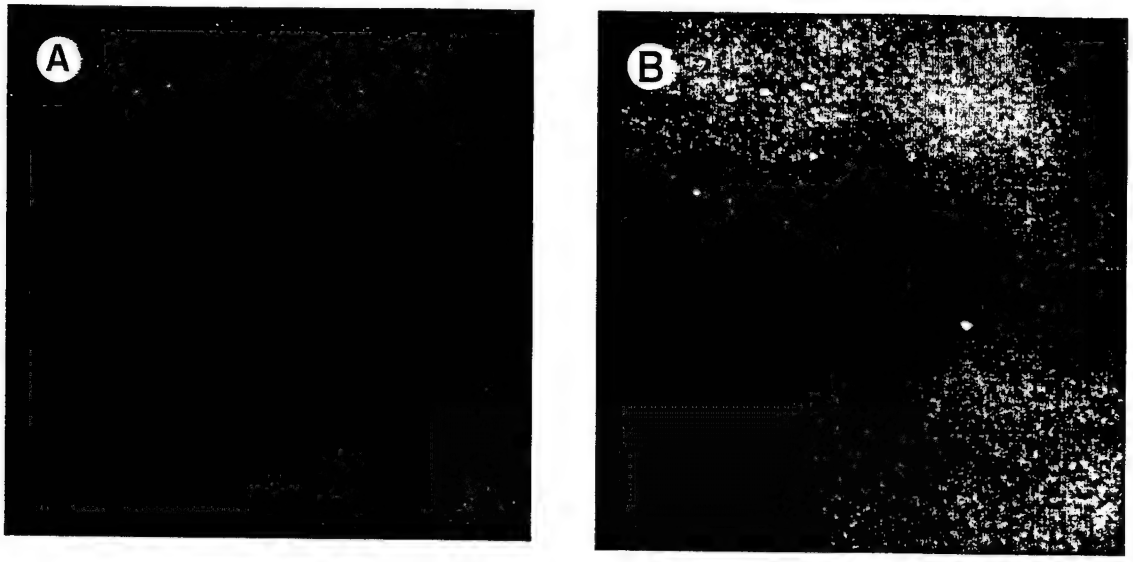

Fig 3. (A) A dense background with a simulated cluster of calcifications embedded in it in the left upper quadrant. The image is enlarged so that the calcifications are readily visualized. (B) The same image as shown in 3A with IW applied. Note how much more obvious the cluster of calcifications appears. The real breast calcification in the right lower quadrant also appears much more obvious with this window.

studies done with two radiologists who specialize in breast imaging (E.D.P. and M.P.B.). In these pilot studies, the two radiologists reviewed dense mammograms with real clinical lesions that were judged to be difficult to visualize using standard screen film mammography. There were seven cases of this type reviewed with 70 combinations of window width and level applied. The radiologists scored each combination of values as showing no change over standard image, improved visibility of the lesion, or worsened visibility of the lesion.

The grid of IW values tested spanned all the likely optimal settings as determined by the pilot work. The IW settings tested were the following: window width 256 with levels 3328, 2456 and 3584; window width 512 with levels 3328, 3456, and 3584; and window width 1024 with levels 3328, 3456, and 3584. The default or unprocessed settings were window width (WW) = 4096, with Level = 2048. There were thus a total of 10 IW settings tested in this experiment.

The digital images were printed onto standard 14 × 17-inch single-emulsion film (3M HNC Laser Film; 3M, St Paul, MN) using a Lumisys Lumicam film printer (Lumisys Inc, Sunnyvale, CA). Each original 50-μm pixel was printed at a spot size of 160 μm, which produced film images 4 × 4 cm, resulting in an enlargement by a factor of three. The radiologist observers in the pilot experiment reported that the magnification did not make the backgrounds unrealistic. Forty images were printed per sheet of film. The images were randomly ordered into an 8 × 5 grid on each sheet of film. Both the film digitizer and film printer were calibrated, and measurements of the relationship between optical density on film and digital units on the computer were determined to generate transfer functions describing the digitizer and film printer. To maintain a linear relationship between the optical densities on the original analog film and the digitally printed film, we calculated a standardization function that provided a linear matching between the digital and printer transfer functions. This standardization function was applied when printing the films to maintain consistency between the original optical densities of the original mammography film and those reproduced on the digitally printed films. The film printer produces films with a constant relationship between an optical density range of 3.35 OD to 0.13 OD, corresponding to a digital input range of 0 to 4095, respectively.

There were 20 observers for the experiment. They were medical students and graduate students from the biomedical engineering and computer science departments. Performance bonus pay was provided. Observers selected the quadrant of the image that they thought contained the cluster of calcifications. All images contained a simulated cluster of calcifications, for a four alternate-forced choice design. Observers were instructed to make their best guess if they could not tell where the simulated lesion was located in the image.

Films were displayed in a dark room on a standard mammography viewbox that was masked to exclude excess light. Observers could move closer to the image, and could use a magnifying glass, if desired. The observers were trained for the task through the use of two sets of images with instructive feedback before actually starting the experiment.

The order of presentation of stimuli was counterbalanced so as to eliminate any effects of learning and fatigue. All 160 possible combinations of processing conditions (10 IW combinations of WW and level), contrast level (four contrasts) and location of the simulated cluster (four quadrants) were used in the experiment. The experiment was designed to have five blocks, in which all 160 combinations appeared. Each observer saw all combinations in each block. All observers completed the experiment. There were 40 backgrounds and four possible rotations of each background, for 160 possible background patterns. For each block, a different background was uniquely assigned to each of the 160 possible processing condition combinations. The assignment was different for each block.

Each observer examined 800 images, for a total of 16,000 stimuli for the whole experiment.

Observers took breaks after each block of images, and more often if necessary. No time limit was imposed on the observation of the images. Typically, the experiment took 2 hours for each observer, divided into two sessions of 60 minutes each. The two sessions were always scheduled on two different days within a week of each other.

DATA ANALYSIS OVERVIEW

Probit models were fit for each subject and enhancement condition using log10 contrast as the predictor. The probability that a subject gets a correct answer is given by the following equation.

$$\Pr(\text{correct}) = 1/4 + (1 - 1/4)\,\phi\,[(x - \mu_{ij})\sigma_i^{-1}]$$

where i indexes subject, and j indexes IW settings. Here ϕ indicates the cumulative Gaussian distribution function. For each subject, this gave a separate location parameter estimate for each IW setting, and a common spread parameter estimate. Assuming a common spread parameter makes sense biologically, as it corresponds to an equal change in log contrast producing an equal change in perception, throughout the visual range. Also, the ¼ arises from the four-choice task.

The location parameter, μ_{ij}, is the mean of the corresponding Gaussian distribution for the ith subject and jth IW setting. Processing conditions that improve detection will cause this parameter to be smaller, and the curve will shift to the left. This occurs because lower contrast levels are required to spot the object. When the processing of the image makes detection harder, higher contrast levels are needed to locate the calcification, and the curve shifts to the right. The values of σ_i, the spread parameter for the ith subject correspond to the slope of the curve. Larger values of σ_i correspond to steep slopes, or greater increase in detection rates per log contrast.

To compare the processing conditions and to examine the effect of window width and level, further analysis was needed. We defined an overall measure to be $\theta_{ij} = \mu_{ij} + \sigma_i$, which corresponds to the log contrast level at which the ith subject viewing the jth IW condition scored 88% correct. We measured the "success" of a processing condition by calculating the difference between the θ score for the unprocessed image and the θ score for the condition for each subject, say $\delta_j = \theta u - \theta_j$, where u is unprocessed. A large positive δ_j score reflects improved performance. It indicates better detection with processed images than with unprocessed images.

Two analyses were performed using this outcome measure. To keep an overall nominal experiment-wise type 1 error rate of .05, a repeated measures analysis of variance was done at the .04 level, with a set nine of *t*-tests at a .01/9 nominal level for each, and hence a .01 level for the whole set.

Repeated measures analysis of variance (ANOVA) allows one to examine the effect of processing conditions and the interactions between window width and level, while accounting for the dependence of measurements taken on the same observer. The repeated measures ANOVA model was fitted, with the δ_j scores as the outcome, and window width and level as the predictors.

RESULTS

The repeated measures ANOVA showed that the interaction between window width and window level was significant at the .04 level (*P* value < .0001, G-G = .729). To examine the nature of this interaction, a series of step-down tests was planned. There was significant interaction between a quadratic trend in window width, and a quadratic trend in window level. Because the quadratic by quadratic interaction was significant, no further tests were examined. A quadratic by quadratic trend means that the surface was curved with respect to both window level and width, and that the shape of the curve differed for fixed values of window width and level (Fig 4).

At the nominal level of .01/9 = .0011, the differences between the default unprocessed condition and the IW conditions were examined. Two settings of intensity windowing processing conditions made finding the calcifications significantly harder, six made the task significantly easier, and one made no significant difference. The settings that made detection easier were window width 1024 with window levels 3328 and 3456 (Table 1, Fig 4).

Average μ_{ij} and σ_i parameters from the best processing condition and the unprocessed condition were used to calculate a typical probit curve. At most, on average, IW processing with settings of window width 1024 and window level 3328 increased the correct detection of calcifications by a maximum of 9%. This is shown in Fig 5.

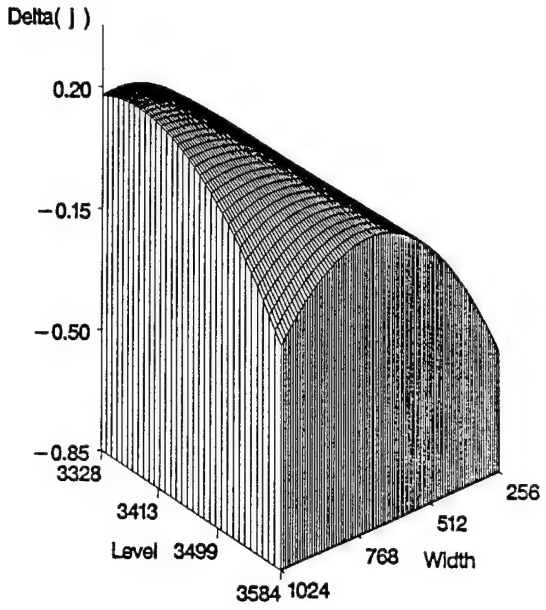

Fig 4. Interpolated predicted values from repeated measures ANOVA: difference in θ value versus window width and level. The peak shows the improved performance due to window width 1024 with window level 3328.

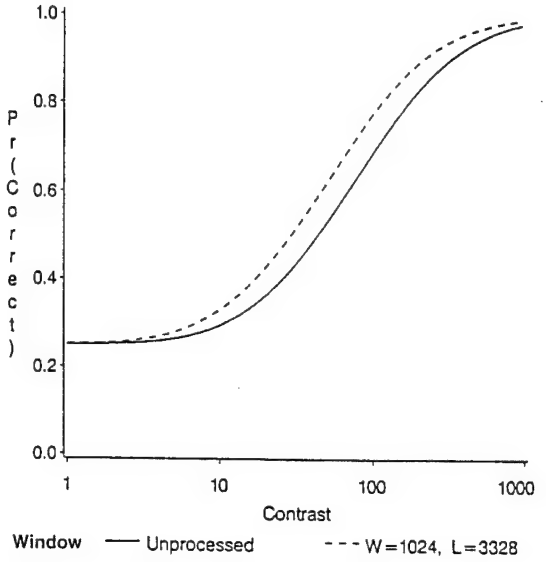

Fig 5. Estimated detection probability for WW of 1024 and level of 3328. The shift in the curve to the left for the processed image reflects improved detection.

DISCUSSION

These results suggest that IW can improve the detection of clustered calcifications on dense mammographic backgrounds, if used properly. Our results also indicate that significant lesion visibility degradation can occur if the window widths and levels are not chosen carefully. We believe that it is important to select the parameters to be applied in the testing of this tool in the clinic based on these types of careful analyses of laboratory studies. Preset intensity windows might then be selected to apply to printed digital mammograms or to mammographic work stations where radiologists might interpret images on line.

Table 1. Mean θ Scores, Difference Scores, and *P* Values for *T* Tests of No Difference

Window Width	Window Level	Mean θ Score	Difference Score	SD	*P* Value
4096	2048	2.46			
256	3328	3.27	−.814	.23	.0001*
256	3456	3.00	−.538	.16	.0001*
256	3584	2.96	−.504	.12	.0001*
512	3328	2.67	−.214	.12	.0001*
512	3456	2.60	−.137	.16	.0012*
512	3584	2.59	−.135	.13	.0002*
1024	3328	2.28	.177	.14	.0001*
1024	3456	2.33	.124	.11	.0001*
1024	3584	2.70	−.246	.10	.0001*

Note: Larger positive difference scores correspond to better performance.

*Significant at the .0011 level.

This work may not predict how this tool will function in a clinical setting. Specifically, graduate student observers and the use of simulated lesions might incorrectly predict the performance of radiologists in detecting real clusters of calcifications in real patients. We have demonstrated previously that graduate student performance at this task parallels the performance of experienced mammographers.[4] The signal-to-noise ratio and the type of image noise present in digital images might vary substantially from digitized mammograms when real full-field digital images are used as the stimuli. Because we have used real clinical images and we have simulated lesions using relatively realistic stimuli, we are optimistic that this image processing algorithm will improve clinical performance. If so, radiologists will be using IW to help them determine whether mammograms of women with dense breasts really do contain calcifications.

Digital mammography is coming to the clinic very soon. It is highly likely that radiologists will want to apply image processing in an attempt to improve their performance in interpreting mammograms. A simple approach to deciding how to view mammograms would be to test every single available algorithm in the clinic on real patients. That would be an expensive and time-consuming process that might have an impact on the care of real

women. It would be preferable, cheaper, and less time-consuming to test this technology in the laboratory before it is tested clinically. The work reported here is intended to help radiologists narrow their choices regarding what might be clinically helpful before expensive clinical tests are undertaken. This project was intended to be a more rigorous exploration of the window widths and levels that might be used clinically in the most challenging areas in the breast, namely the dense parts.

Furthermore, specific IW values depend on the calibration of the instrumentation used for digitization or acquisition, and the patient being imaged. IW values are not standardized and therefore may not directly translate from system to system. That is, the IW values reported on here may not be the correct ones for a different system. However, this experiment showed that there are IW values that can significantly improve detectability of calcifications as well as IW values that substantially degrade lesion visibility. With the advent of full-field digital mammography, and with the standardization of data acquisition, IW values could also be standardized across systems.

This experiment does not address how IW would affect the appearance of fatty areas of the breast, and the detection of calcifications in those parts. We would not want to view a mammogram solely with an algorithm applied that degrades performance in areas where sensitivity is currently quite high. If this algorithm is useful in dense areas, it could potentially be applied selectively to only the dense parts of the breast. Alternatively, it could be used as an adjunct with the image viewed in a standard format, and then with the calcification window width and level applied.

Our experiments to date cannot estimate the frequency of false positives when IW would be used clinically. Many alternate forced choice tests yield proportion correct as the primary outcome. Macmillan and Creelman describe methods for converting proportion correct in this setting to a value for d′, the sensitivity parameter of an ROC analysis.[17] Given the characteristics of the study design, subjects, and training, we believe that superior proportion correct will translate into superior d′. Of course, this must be proven in a true clinical setting with ROC analysis.

REFERENCES

1. Homer MJ: Mammographic Interpretation: A practical approach. New York, NY, McGraw Hill, 1991, pp 4-5
2. Rosenman J, Roe CA, Cromartie R, et al: Portal Film enhancement: Technique and clinical utility. Int J Radiat Oncol Biol Physics 25:333-338, 1993
3. Shtern F: Digital mammography and related technologies: A perspective from the National Cancer Institute. Radiology 183:629-630, 1992
4. Puff DT, Pisano ED, Muller KE, et al: A method for determination of optimal image enhancement for the detection of mammographic abnormalities. J Dig Imaging 7:161-171, 1994
5. Pisano ED, Chandramouli J, Hemminger BM, et al: Utility of Intensity Windowing in Improved Detection of Simulated Masses in Mammograms of Dense Breasts. Presented at the Radiologic Society of North America Meeting. Chicago, IL, November 27, 1995
6. McSweeney MB, Sprawls P, Egan RL: Enhanced Image Mammography. AJR 140:9-14, 1983
7. Smathers RL, Bush E, Drace J, et al: Mammographic microcalcifications: Detection with xerography, screen film, and digitized film display. Radiology 159:673-677, 1986
8. Chan HP, Doi K, Galhorta S, et al: Image Feature analysis and computer-aided diagnosis in digital radiography: I. Automated detected of microcalcifications in mammography. Med Phys 14:538-547, 1987
9. Chan HP, Vyborny CJ, MacMahon H, et al: Digital mammography ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. Investigative Radiol 22:581-589, 1987
10. Hale DA, Cook JF, Baniqued Z, et al: Selective Digital Enhancement of Conventional Film Mammography. J Surg Oncol 55:42-46, 1994
11. Yin F, Giger ML, Vyborny CJ, et al: Comparison of Bilateral-Subtraction and Single-Image Processing Techniques in the Computerized Detection of Mammographic Masses. Investigative Radiol 28:473-781, 1993
12. Yin F, Giger M, Doi K, et al: Computerized detection of masses in digital mammograms: Analysis of Bilateral Subtraction Images. Med Phys 18:955-963, 1991
13. Pizer SM: Psychovisual issues in the display of medical images. KH Hoehne, ed, Pictorial Information Systems in Medicine. Berlin, Springer-Verlag, 1985, pp 211-234
14. Revesz G, Kundel HL, Graber MD: The influence of structured noise on the detection of radiologic abnormalities. Investigative Radiol 9:479-486, 1974
15. Kundel HL, Reversz G: Lesion conspicuity, structured noise and fil reader error. AJR 126:1233-1238, 1976
16. Revesz G, Kundel HL: Psychophysical studies of detection errors in chest radiology. Radiology 128:559-562, 1977
17. MacMillan NA, Creelman CD: Detection theory: A user guide. Cambridge, England, Cambridge, 1991, pp 135-136

Appendix A.3

Hemminger BM, Dillon A, Johnston RE, Muller K, Pisano ED, Deluca M. Evaluation of the effect of display luminance on the feature detection of simulated masses in mammograms. SPIE Medical Imaging 1997;3036:12.

Evaluation of the Effect of Display Luminance on the Feature Detection Rates of Masses in Mammograms

Bradley M. Hemminger*, Alan Dillon#, R. Eugene Johnston*,

*Department of Radiology, University of North Carolina,
#Department of Biomedical Engineering, University of North Carolina,
Chapel Hill NC 27599
bmh@rad.unc.edu

ABSTRACT

Purpose To determine the interaction of the luminance range of the display system with the feature detection rate for detecting simulated masses in mammograms.

Methods Simulated masses were embedded in cropped 512x512 portions of mammograms digitized at 50 micron pixels, 12 bits deep. The masses were embedded in one of four quadrants in the image. An observer experiment was conducted where the observer's task was to determine in which quadrant the mass is located. The key variables involved in each trial included the position of the mass, the contrast level of the mass, and the luminance of the display. The contrast of the mass with respect to the background was fixed to one of four selected contrast levels. The digital images were printed to film, and displayed on a mammography lightbox. The display luminance was controlled by the placing neutral density films between the laser printed films of mammographic backgrounds and the lightbox. The resulting luminances examined in this study ranged from a maximum of 10 ftL to 600 ftL. Twenty observers viewed 20 different combinations of the 5 neutral density filters with the 4 contrast levels, for a total of 400 observations per observer, and 8000 observations overall.

Results An ANOVA analysis showed that there was no statistically significant correlation between the luminance range of the display and the feature detection rate of the simulated masses in mammograms. None of the luminance display ranges performed better than any of the others.

Key Words: Image Display, Luminance, Masses, Feature Detection, Display System Characteristics, Mammography, Observer Studies.

SPIE Vol. 3036 • 0277-786X/97/$10.00

2. BACKGROUND AND SIGNIFICANCE

In the past, the medium of film has served as both the storage and the display media for medical imaging. Today, with the advent of digital modalities for most every Radiology examination, and the convenient transmission of digital medical image data via the DICOM communications standard, a decoupling of image storage and image display has occurred. This decoupling is significant, in that images can now be processed prior to their display, and the display of images need not be dependent on limitations of the acquisition and storage systems. As a result, it is important to study the characteristics necessary for medical image display, once the image storage component is separated from the image display component. The specific question addressed in this research is what maximum luminance level is necessary for medical image display.

We chose to evaluate mammography because it has the strictest requirements for luminance range of radiologic medical image display devices. Specifically the ACR recommends 1000ftL luminance lightboxes for the display of analog film-screen mammography films. This requirement is due a number of factors, including the film characteristic curve, limitations inherent in the analog film-screen acquisition techniques, and ambient light of the viewing setting. Now that the image data can be acquired digitally, however, the luminance range of the display device can be determined independently from the acquisition parameters. We would like to determine whether display systems with smaller maximum luminances than the currently proscribed 1000 ftL requirement can perform as well. If they do, then softcopy (video) displays may be satisfactory for mammography image presentation. Additionally, these results should be similar for other, less demanding modalities. This study attempts to determine the effect of the luminance range of display systems on the feature detection rate of masses in mammograms. Masses were chosen because this is similar to many radiology detection tasks (masses in lungs on chest Xray, nodules on chest CT, etc.). The maximum luminances evaluated in the experiment were chosen to match those commercially available for video display systems and mammography lightboxes.

3. MATERIALS AND METHODS

The experimental paradigm used is based on the model we have previously described for evaluating feature detection and contrast enhancement for medical image display.[1,2,3] It allows for the laboratory testing of a range of display parameters (in this case, the luminance range of the display system). Simulated masses were embedded in cropped 512x512 portions of mammograms digitized at 50 micron pixels, 12 bits deep. The masses were embedded in one of four quadrants in the image. An observer experiment was conducted where the observer's task was to determine in which quadrant the mass is located. The key variables involved in each trial included the position of the mass, the contrast level of the mass, and the luminance of the display. The contrast of the mass with respect to the background was fixed to one of four selected contrast levels. The digital images were printed to film, and displayed on a mammography lightbox. The display luminance was controlled by the placing neutral density films between the laser printed films of mammographic backgrounds and the lightbox. The resulting luminances examined in this study ranged from a maximum of 10 ftL to 600 ftL. Twenty observers viewed 20 different combinations of the 5 neutral density filters with the 4 contrast levels, for a total of 400 observations per observer, and 8000 observations overall.

Mammographic Backgrounds

The 80 background images of 512x512 pixels each were taken from clinical mammograms that had been digitized using a Lumiscan digitizer (Lumisys, Inc., Sunnyvale, CA) with a 50

micron sample size and 12 bits of intensity data per sample. The images were selected so as to provide an even distribution of density distributions across density range of breast tissue on clinical mammograms. The mammograms were known to be normal by virtue of 3 years of clinical and mammographic follow-up. They were selected by a radiologist expert in breast imaging from digitized film screen craniocaudal or mediolateral oblique mammograms.

The gray scale values for the mammographic backgrounds are assigned the values recorded by the Lumisys digitizer. The digitizer assigns digital values in the range 495-4095 representing an optical density range of 3.68 - 0.02. The digitizer produces digitized grey values that map one to one with OD values, i.e., the same OD value on film will produce the same grey level when digitized.

Mammographic Mass Stimuli

Mammographic masses were simulated using a locally developed program. A circle of diameter of 90 pixels was generated. When printed on film the mass was 7.2 mm in diameter, and 1' of viewing angle at the average viewing distance of 40cm (about 16"). The circle was gaussian blurred (frequency standard deviation of 0.2) to appear similar to masses presenting on clinical mammograms. Simulated masses were used instead of real features so that we could have precise control over the structure location, and structure to background contrast of the masses. While the simulated masses were not perfectly realistic, our mammographers confirmed that they did possess the same scale and similar spatial characteristics to actual masses seen at mammography.

Contrast

The contrast of the mass to background surround was defined as the luminance ratio ($\Delta L/L$) where ΔL was the luminance of the background surround with the target inserted minus the luminance of the background surround without the target inserted. L is the luminance of the background surround. Several different choices exist for the area under which the mean background surround value could be calculated. Some common choices are depicted in figure 1. While we believe the definition best matched to visual perception would be one that takes into account the structure of the background surround, there are presently no established techniques for this option. We chose for this experiment to use the area just under the inserted target mass feature. We investigated whether choosing a different size area for calculating the mean of the background surround would have effected our calculation of contrast values. Analysis of randomly inserting 1000 target masses into each of the 80 mammographic background images used in the experiment and calculating the resulting mean background surround value, showed that using increasingly larger circles for the background surround area, up to the size of the mammographic background image, did not significantly change the mean digital driving level used for the surround, as compared to the size of the smallest contrast steps used in the experiment. The standard deviation, however, as might be expected, did increase with the larger circles due to the larger inclusion criteria. Thus, using larger diameter circles could possibly reduce sensitivity in measuring the detection rate due to increased variance in calculation of the mean of the background surround. This result supported our decision to use the surround area equal to the area under the target (i.e. a smaller diameter circle).

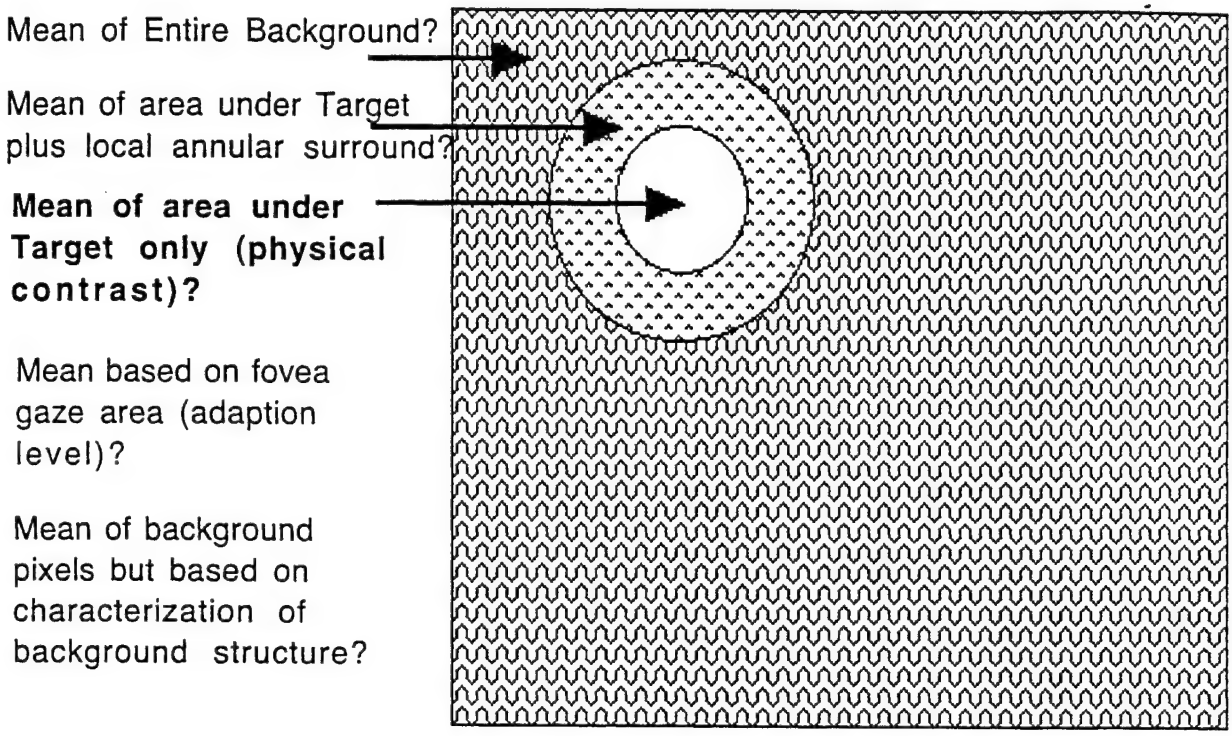

Figure 1. A representation of mammographic background, with white interior circle depicting the mass target insertion area. The surrounding annular ring shows an example of a larger including circle for which the mean could be calculated. Mean values are calculated in digital driving levels of the computer display device, which can be translated into luminance values. Five different methods of calculating the surround value for the contrast definition are given.

The masses were embedded at one of four different contrast levels by pixel-wise addition of the structure and background images. The contrast levels were equally spaced in perceived brightness relative to mean luminance of the background surround area. To calculate the contrast we first calculated the mean DDL in the area of the background where the target would be placed. Then we calculated from this the luminance that would be produced in the experimental setting when this film was placed on our mammography lightbox based on calibration measurements of the printed films on this lightbox. From this surround luminance value, we calculated the luminance value that the target stimulus (mass) should be in order to give us the desired contrast level, and then performed the reverse calculation to determine the DDL values for the mass.

Contrast levels were chosen to provide appropriate calculation of the probit curve. Initial choices of contrast level values were estimated from our prior work. Then we piloted the experiment with 3 observers on a separate set of cropped background images similar to the study ones. Sufficient numbers of trials were used to obtain reasonable estimates of contrast thresholds. The pilot experiments were continued until the chosen contrast levels were appropriately spaced to properly define the probit curve. For this experiment we repeated the pilot three times, each time using 32 or 64 trials repeated with each neutral density film, and with 3 observers. The final contrast levels chosen were contrast values of 4%, 10%, 16% and 22%. This corresponded to percent correct detection rates of 30%, 50%, 80%, 95%, respectively.

Experimental Presentation

The digital images were printed onto standard 14X17 inch single emulsion film (3M HNC Laser Film, 3M, St Paul, MN) using a Lumisys Lumicam film printer (Lumisys Inc, Sunnyvale, CA). Each original 50 micron pixel was printed at a spot size of 80 microns, which produced film images enlarged by a factor of 1.6, approximately 4x4 centimeters in size. Radiologist observers in the previous experiments using this same paradigm reported that they felt this magnification did not make the backgrounds unrealistic.[3] Thirty-two cropped backgrounds were printed per sheet of film. The backgrounds were randomly ordered into an 8X4 grid on each sheet of film. The 8x4 grid was chosen because the mammography lightbox was uniform in luminance only over the central portion of the lightbox, which corresponded to the 32cmx16cm area covered by the 8x4 image grid. The mean luminance of the film test image displayed on the mammography lightbox without any filters was 18 ftL, 26 ftL, and 19 ftL, respectively, for the three film test images in the experiment.

Both the film digitizer and film printer were calibrated, and measurements of the relationship between optical density on film and digital units on the computer were determined in order to generate transfer functions describing the digitizer and film printer.[2] In order to maintain a linear relationship between the optical densities on the original analog film and the digitally printed film, we calculated a standardization function that provided a linear matching between the digitizer and printer transfer function curves, so that, for example, an OD in the 15 percentile on the digitizer curve would map to the OD on the 15 percentile on the film printer curve. This standardization function was applied to the mammographic image backgrounds so that the printed films would maintain a consistent proportional relationship between the original optical densities of the original mammography film and those reproduced on the digitally printed films. The film printer produces films with a constant relationship between an optical density range of 3.62 OD to 0.13 OD, corresponding to a digital input range of 0 to 4095, respectively.

We choose to use neutral density films to control the luminance of the display for consistency, and because of the inherent maximum luminance capability of the lightbox. If we had used a video display system such as a CRT, we would not have be able to reproduce the high luminance levels of lightboxes. Additionally, we would have had to sacrifice contrast resolution (number of grey levels utilized) in order to drive the monitor at reduced luminance ranges in a consistent fashion. Similarly, if we produced films with different luminance ranges, we would have had to decrease the contrast resolution because of only using part of the grey scale range of the display device. It would have additionally caused us to produce multiple films for different luminance values. Since variables in the film printing process could cause differences between films depicting the same contrast levels, producing multiple films for different luminance values might have added a confounding variable to our analysis. For the above reasons, we chose to print a single version of the test images, and use neutral density films instead to modify the luminance of the display. The neutral density films were created using the same Lumicam laser printer used to print the mammographic backgrounds. Uniform flat field films of constant density were produced for the neutral density backgrounds. We also evaluated photographically producing the neutral density filters, but found the variance of OD to be larger for the photographically produced films than for the laser printed neutral density films. We scanned the neutral density films on our Lumisys scanner to check their uniformity. The means and standard deviations of the digitized neutral density films are shown in table 2.

Neutral Density Filter	Mean	STDDEV
10	2230	51
20	2490	27
30	2672	24
200	3539	10
600	3971	6

Table 2. Digital Driving levels of digitized neutral density films. These measurements were used to determine how much noise the neutral density films added to the overall experimental process. Films produced optically had a significantly higher STDDEV.

The resulting luminance levels on the lightbox were exactly controlled by using a voltage regulator inline with the lightbox to adjust the luminance level of the lightbox output up or down. By measuring the luminance output of the lightbox through the neutral density film with a photometer before each experimental session, we could tune the voltage regulator to set the transmitted luminance to exactly the desired output level. This allowed us to consistently maintain the experimental luminance settings for the neutral density films throughout the experiment. The maximum luminance levels in the experiment were chosen to match common commercially available luminance levels for video display systems and mammography lightboxes. The values selected were mammography lightbox (600ftl, see explanation below as to why different from 1000 ftL), high brightness CRT (200 ftL), average workstation monitor (30 ftL), and low end personal computers or hardcopy displays (20 ftL and 10 ftL). The values chosen for the luminance levels are the values as measured by a photometer through one of the five neutral density films combined with either a 0 DDL test film (low end of luminance range) or a 4095 DDL test film (the high end luminance range). These two test films represented the darkest and brightest images possible on the display system with laser printed film. For instance, the high end of the brightest luminance consisted of a clear film undeveloped as the neutral density film, and on top of that a test film produced by the laser printer using the maximum digital driving level of 4095 to produce a uniform flat field. Our mammography lightbox actually produced 790 ftL rather than the expected 1000 ftL. Thus, the highest maximum luminance produced on our mammography display using neutral density films was 600ftL, as measured through the clear neutral density film. Table 3 shows the optical densities of the five neutral density films, the 0 DDL test film, and the 4095 test films, and lists the measured luminances used in this experiment (i.e. what is measured transmitted through the neutral density films and test films on the mammography lightbox).

	0 DDL Test Film OD = 3.62	4095 DDL Test Film OD = 0.13	Range (Max/Min)
ND0 (OD = 1.81)	0.0016 ftL	7.3 ftL	4563
ND1 (OD = 1.56)	0.0031 ftL	14.3 ftL	4613
ND2 (OD = 1.39)	0.0048 ftL	21.8 ftL	4542
ND3 (OD = 0.55)	0.0341 ftL	146.0 ftL	4282
ND4 (OD = 0.13)	0.1120 ftL	457.0 ftL	4080

Table 3. Values show transmitted luminance from lightbox through different neutral density films and min and max test films (DDL 0 and DDL 4095 on laser film printer). Maximum luminance of lightbox without any films is 790 ftL. Rightmost column shows the calculated dynamic range of the display condition (maximum luminance divided by minimum luminance).

The experiment was conducted in our experimental laboratory, which is controlled for light, sound, and other distractions. Room light was 0.043 (day) to 0.0065 (night) lux with no images displayed, and an average of 0.225 lux, 0.376 lux, 0.671 lux, 3.98 lux, 10.63 lux, when experimental films were displayed using the neutral density filters of 10ftL, 20ftL, 30ftL, 200ftL,

600ftL, respectively. Films were displayed on a standard mammography viewbox that was masked to exclude excess light. Observers were free to move, and could use a standard mammography magnifying glass, if desired. Average viewing distance was 16". Observers were dark adapted to the light levels of the experiment for 10 minutes prior to any readings. The neutral density film was placed first on the lightbox. The mammography test film was placed directly on top of the neutral density film.

Observation Task

There were 20 observers for the experiment. They were medical students and graduate students from the University of North Carolina. Performance bonus pay was used to encourage optimal observer performance. Observers selected the quadrant of the image that they thought contained the mass. All images contained a simulated mass, for a 4 Alternative-Forced Choice design. Observers were instructed to make their best guess if they could not tell where the simulated lesion was located in the image.

Prior to beginning the experiment, observers were trained for the task through the use of two films each with 64 images. The first 32 images contained easy (high contrast cases), and the second 32 images contained cases with the contrast matching the levels used in the experiment. An answer sheet overlay provided feedback indicating the correct location of the mass on each image.

The order of presentation of stimuli was counterbalanced so as to eliminate any effects of learning and fatigue. Observers were encouraged to take breaks if needed. Observers were dark adapted to the room upon re-entry. All observers completed the experiment. Each observer examined 80 different images, with the 5 neutral density combinations, for a total 400 images per observer, and a total of 8000 stimuli for all observers for the whole experiment.

Observers took a break at the half way point during the study, and more often if necessary. No time limit was imposed on the observation of the images. Typically, the experiment took 2 hours for each observer, divided into two sessions of 60 minutes each, with a 5 minute break in between sessions.

4. DATA ANALYSIS

Probit models were fit for each subject and display luminance using log10 contrast as the predictor. The probability that a subject gets a correct answer is given by the following equation.

$$\Pr(\text{correct}) = 1/4 + (1 - 1/4)\,\Phi\left[(X - \mu_{ij})\,1/\sigma_i\right]$$

where i indexes subject and j indexes luminance settings. Here Φ indicates the cumulative Gaussian distribution function. For each subject, this gave a separate location parameter estimate for each luminance setting, and a common spread parameter estimate. A common spread parameter is assumed, since this corresponds with what is known biologically about the human visual system (i.e. it corresponds to an equal change in log contrast producing an equal change in perception throughout the visual response range corresponding to the luminance range of this experiment). The 1/4 arises from the 4 AFC task.

The location parameter, μ_{ij}, is the mean of the corresponding Gaussian distribution for the ith subject and the jth luminance setting. Display luminance conditions that improve detection will cause this parameter to be smaller, and the curve will shift to the left. This occurs because lower contrast levels are required to spot the object. When the display condition makes detection harder, higher contrast levels are needed to locate the mass, and the curve shifts to the right. The values of

σ_i, the spread parameter for the ith subject, correspond to the slope of the curve. Smaller values of σ_i correspond steeper slopes, or greater increases in detection rates per log contrast.

Repeated measures of analysis of variance (ANOVA) allows one to examine the effect of display luminance level, while accounting for the dependence of measurements taken on the same observer. The repeated measures ANOVA model was fitted, with the σj scores as the outcome. The log10 contrast was the predictor for this model.

To compare the processing conditions and to examine the effect of luminance, further analysis was needed. We defined the overall measure to be $\theta ij = \mu_{ij} + \sigma_i$, which corresponds to the log contrast level at which the ith subject viewing the jth luminance condition scored 88% correct. We measured the effect of display luminance condition by calculating the delta (σj) difference between the θ score for the display condition of 600 (reference standard of mammography lightbox) and the θ score for each of the other display luminance conditions, for each subject in this study. A larger positive σj score reflects improved detection, which indicates a more negative θj value. This would indicate better detection with other display luminance conditions than with the standard display luminance condition.

Two analyses were performed using this outcome measure. In order to keep a nominal overall type 1 error rate of 0.05 for experiment, a first repeated measures analysis of variance was done at the 0.04 level, and second set of 4 T-tests was performed at a 0.01 level (0.04 + 0.01 = 0.05). Since there were 4 T-tests, each was performed at 0.01/4 = 0.0025 level. A total of 20 subjects were tested.

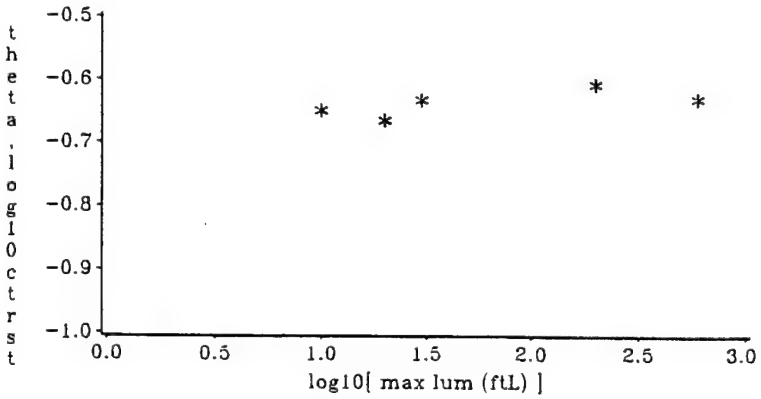

Figure 2. Shows the mean theta values for each level of maximum luminance display condition (luminance is expressed as log10). Rightmost point is 600 ftL condition, and leftmost point is 10 ftL condition. Values closer to the bottom indicate lower contrast thresholds where the observers were more sensitive.

The repeated measures analysis of variance revealed that display luminance condition did not significantly effect the threshold for the detection of masses at the 0.04 level (p-value = 0.0832, Geiser-Greenhouse epilson^ = .6261, df = 4). These results are shown in Figure 2, which depicts the mean θ values for each display luminance condition.

The second analysis, the series of planned step-down tests was implemented at the nominal level of 0.01/4 = 0.0025. The differences between the standard luminance condition and the remaining conditions were examined. None of the P-values were less than 0.0025, and thus none of the display luminance conditions made a significant difference in correctly locating the masses. These results are seen in table 4, which gives the summary statistics for σj at different luminance conditions.

	Mean	Std Deviation	P Value
δ_{10_600}	+ .0203	.1055	0.3998
δ_{20_600}	+ .0358	.0613	0.0173
δ_{30_600}	+ .0038	.0694	0.8094
δ_{200_600}	- .0229	.1034	0.3339

Table 4. Summary statistics for δ at different display luminance level differences, where (δx_y represents difference between scores for display luminance conditions x and y)

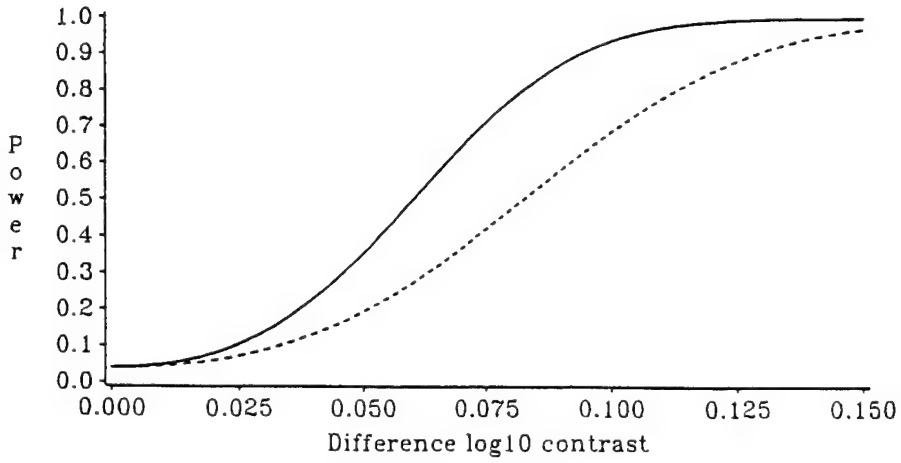

Figure 3. This figure shows the power curve for this experiment. The solid line is estimated power, and the dashed line is the 95% lower bound confidence interval.

Finally a retrospective power analysis was computed. Figure 3 shows the power required to detect a difference in log10 of contrast for a repeated measures of analysis of variance at the 0.04 level The solid line is the estimated power, and the dashed line is the exact 95% one-sided (lower bound) confidence band on power.

5. DISCUSSION

Digital mammography is already beginning to appear in the clinic. It is highly likely that several methods of displaying digital mammograms will be available. It is important to characterize what effects the display system will have on the radiologists' clinical performance. These results suggest the display luminance of the display system is not a significant factor affecting the detection rate of simulated masses inserted in mammographic backgrounds. Vision theory would predict this for uniform backgrounds for this luminance range where Webber's law holds (the value of $\Delta L/L$ is constant). This result validates this for mammographic backgrounds and mass targets. It suggests that lower luminance display systems may function just as well for detection tasks in radiology. Specifically, the option of lower luminance video displays may be a viable option.

The biggest caveat is that the lower luminance levels for which an effect was not found (10ftL to 30ftL) would probably not perform as well under actual clinical conditions. This is because most clinical reading rooms have too much ambient light (overhead fluorescent lights) and glare (from surrounding lightboxes). These light levels are known to cause the contrast thresholds to be larger for the lower luminance display systems. Thus, the result of no significant differences for those display luminance levels may not hold for actual clinical conditions, unless the working environments are changed. Under such clinical conditions these results still suggest that the brighter CRT monitors that are currently commercially available should provide sufficient range for mammographic image presentation, and likely for most other radiological image displays as well, while not being compromised by room lighting conditions.

An important side issue of this talk is the discussion of what contrast definition to use. This is an area requiring further work, especially in the area of background structure and texture based surround luminance measures. Standardization of measures of contrast for non-uniform backgrounds would be of significant help in allowing comparison across different research results.

6. FUTURE WORK

Important future work would be to extend these results to other radiological backgrounds and feature targets, and to test under clinical room lighting conditions. We also plan to conduct similar studies on video displays.

7. ACKNOWLEDGMENTS

This work was supported in part by NIH grants P01-CA47982, R01-CA60193, and R01-CA44060.

8. REFERENCES

1. Puff DT, Pisano ED, Muller KE, Johnston RE, Hemminger BM, Burbeck CA,

McLelland R, Pizer SM, "A Method for Determination of Optimal Image Enhancement for the Detection of Mammographic Abnormalities", Journal of Digital Imaging, Nov 1994.

2. Hemminger BM, Johnston RE, Rolland JR, Muller KE, "Perceptual Linearization of Video Display Monitors for Medical Image Presentation", Proceedings of Medical Imaging 1994: Image Capture, Formatting and Display, SPIE Vol 2164, 1994.

3. Pisano EP, Chandramouli J, Hemminger BM, DeLuca M, Blueck D, Johnston RE, Kuller K, Braeuning MP, Pizer SM, "Does Intensity Windowing Improve the Detectation of Simulated Microcalcifications in Dense Mammograms?", JDI, accepted for May 1997.

Appendix A.4

Pisano ED, Zong S, Hemminger BM, DeLuca M, Johnston RE, Muller K, Braeuning MP, Pizer S. Contrast Limited Adaptive Histogram Equalization Image Processing to Improve the Detection of Simulated Spiculations in Dense Mammograms. Journal of Digital Imaging. 1998; 11(4): 193-200.

Contrast Limited Adaptive Histogram Equalization Image Processing To Improve the Detection of Simulated Spiculations in Dense Mammograms

Etta D. Pisano, MD^*

Shuquan Zong, MS+

Bradley M. Hemminger, MS^

Marla DeLuca, MS@

R. Eugene Johnston, Ph.D.^

Keith Muller, Ph.D@

M. Patricia Braeuning, MD^*

Stephen M. Pizer, Ph.D.^*`

From the Departments of Radiology^, Computer Science`, Biomedical Engineering° and Biostatistics".

The University of North Carolina-Chapel Hill.

School of Medicine, School of Public Health and College of Arts and Sciences.

Member, UNC-Lineberger Comprehensive Cancer Center*.

Supported by NIH PO1-CA 47982, NIH RO1-65583 and DOD DAMD 17-94-J-4345.

Corresponding author:

Etta D. Pisano, MD

Department of Radiology

503 Old Infirmary Building CB# 7510

Chapel Hill, NC 27599-7510

(919) 966-6957

Fax: (919) 966-0817

Contrasted Limited Adaptive Histogram Equalization Image Processing To Improve the Detection of Simulated Spiculations in Dense Mammograms

Abstract

Purpose:

To determine whether Contrast Limited Adaptive Histogram Equalization (CLAHE) improves detection of simulated spiculations in dense mammograms.

Methods:

Lines simulating the appearance of spiculations, a common marker of malignancy when visualized with masses, were embedded in dense mammograms digitized at 50 micron pixels, 12 bits deep. Film images with no CLAHE applied were compared to film images with nine different combinations of clip levels and region sizes applied. A simulated spiculation was embedded in a background of dense breast tissue, with the orientation of the spiculation varied. The key variables involved in each trial included the orientation of the spiculation, contrast level of the spiculation and the CLAHE settings applied to the image. Combining the 10 CLAHE conditions, 4 contrast levels and 4 orientations gave 160 combinations. The trials were constructed by pairing 160 combinations of key variables with 40 backgrounds. Twenty student observers were asked to detect the orientation of the spiculation in the image.

Results:

There was a statistically significant improvement in detection performance for spiculations

with CLAHE over unenhanced images when the region size was set at 32 with a clip level of 2, and when the region size was set at 32 with a clip level of 4.

Major Conclusion:

The selected CLAHE settings should be tested in the clinic with digital mammograms to determine whether detection of spiculations associated with masses detected at mammography can be improved.

Key Words: Mammography, Image Processing, Contrast Limited Adaptive Histogram Equalization, Observer Studies, Breast Cancer, Spiculations

Background and Significance

Approximately 10-15% of palpable malignancies are not visible mammographically (1). It is highly likely that many nonpalpable cancers are also not visible with current technology. Digital mammography might allow for greater contrast and improved detection of small and early tumors over standard film screen technology, especially if image processing is used to improve image contrast (2-5).

We have previously published two papers reporting laboratory results that show improved performance by students in finding simulated masses and simulated clustered calcifications embedded in dense mammographic background when Intensity Windowing is applied compared to their performance when viewing non-windowed images (6, 7). The methods used in those experiments were based on methods reported in a previous paper (8) in which we demonstrated that detection performance with the application of Contrast Limited Adaptive Equalization (CLAHE) to digitized mammograms is parallel for radiologists and student observers (8). Using the same experimental paradigm, we report here that Contrast Limited Adaptive Histogram Equalization (CLAHE) can improve the detection of simulated spiculations in dense mammograms in a laboratory setting.

Many investigators have studied the use of image processing techniques in digitized mammograms. McSweeney attempted to improve the visibility of calcifications by using edge detection for small objects, but gave no clinical results (9). Smathers improved the

visibility of small objects in images by intensity band-filtering (10). Chan used unsharp-masking to reduce image noise to improve detection of clustered calcifications (11). Chan, Hale, and Yin have tested other image processing methods on digitized mammograms with variable results (12-15). Kallergi et al. have demonstrated improved radiologist performance in detecting clustered calcifications in wavelet- processed digital mammograms vs. unenhanced digital mammograms (16).

Contrast enhancement methods are not designed to increase or supplement the inherent structural information in an image, but rather to improve the image contrast and theoretically to enhance particular characteristics. Contrast Limited Adaptive Histogram Equalization is an adaptive contrast enhancement method. It is based on adaptive histogram equalization (AHE) [17], where the histogram is calculated for the contextual region of a pixel. The pixel's intensity is thus transformed to a value within the display range proportional to the pixel intensity's rank in the local intensity histogram. CLAHE [18] is a refinement of AHE where the enhancement calculation is modified by imposing a user-specified maximum, i.e., clip level, to the height of the local histogram, and thus on the maximum contrast enhancement factor. The enhancement is thereby reduced in very uniform areas of the image, which prevents overenhancement of noise and reduces the edge-shadowing effect of unlimited AHE. The size of the pixels' contextual region and the clip level of the histogram are the parameters of CLAHE (18).

The experiments described in this paper were performed to determine whether Contrast

Limited Adaptive Histogram Equalization could improve the detection of simulated spiculations in dense mammograms in a laboratory setting. While the scope of this paper is limited to the evaluation of observer performance with respect to the contrast of the simulated spiculations to background using our established experimental paradigm, it may be interesting for follow-up work to evaluate these results with respect to measures proposed by other authors, such as the conspicuity measure proposed by Revesz and Kundel (19-21).

Materials and Methods

The experimental paradigm used here is based on the model we have previously described and allows for the laboratory testing of a range of parameter values (in this case, region size and clip level). (4). The experimental subject is shown a series of test images that consist of an area of a dense mammogram with a simulated spiculation embedded in the image in one of four orientations. The observer's task is to determine in which orientation the line is located. The test images are displayed in both the processed and unprocessed format, and the contrast of the object against the background is varied from quite easy to detect to impossible to detect.

A computer program randomly selected one of 40 background images and rotated that background to one of four orientations. The 40 backgrounds images of 512X512 pixels each were taken from actual mammograms that had been digitized using a Lumiscan

digitizer (Lumisys, Inc., Sunnyvale, CA) with a 50 micron sample size and 12 bits of intensity data per sample. The images were selected from relatively dense parts of the mammograms that were known to be normal by virtue of 3 years of clinical and mammographic follow-up. They were selected by a radiologist expert in breast imaging from digitized film screen craniocaudal or mediolateral oblique mammograms.

A grey scale value for each pixel of the digitized mammographic background is assigned a value recorded by the Lumisys digitizer. The digitizer assigns digital values in the range 495-4095 representing an optical density (OD) range of 3.43-0.08. The digitizer produces digitized grey values that map one to one with OD values, i.e., the same OD value on film will produce the same grey level.

The 40 different dense backgrounds were utilized. A phantom feature, the simulated spiculation, was then added into the background. The image was then processed with CLAHE to yield the test stimulus.

A spiculation was simulated using a 13-18 mm. long line, 160 microns wide. Simulated spiculations were used instead of real features so that we could have precise control over the structure location, orientation and structure to background contrast of the pseudolesions. To more realistically simulate spiculated masses would have required using multiple pixels per spiculation, for instance a 2 pixel wide or 3 pixel wide matrix. Because of limitations of our printer which had a spot size of 160 microns per pixel, the use of a

wider spiculation would have unrealistically enlarged the simulated spiculations. Thus we limited our simulated lesions to single pixel wide areas, and varied only the contrast of the spiculation. As a result, the simulated spiculations were not entirely realistic, but they did possess the same scale and similar spatial characteristics to actual spiculations seen at mammography.

The intensity difference of the spiculations from background was defined as the grey level of the digital spiculations before addition to the background. The spiculations were then embedded at four different orientations with four different intensity levels equally spaced in perceived brightness relative to background by pixel-wise addition of the structure and background images. Figures 1 and 2b show an example of a simulated spiculation. Figure 2a shows a set of real spiculations within a specimen radiograph for comparison.

A three by three (3X3) grid of appropriate region size and clip level parameter settings was selected based on the results of pilot preference studies done with two radiologists who specialize in breast imaging (EDP and MPB). In these pilot studies, the two radiologists reviewed dense mammograms with real clinical lesions that were judged to be difficult to visualize using standard screen film mammography. There were 7 cases of this type reviewed with 70 combinations of region size and clip level applied. The radiologists scored each combination of values as showing no change over standard image, improved visibility of the lesion, or worsened visibility of the lesion.

The grid of CLAHE values tested spanned all the likely optimal settings as determined by the pilot work. The CLAHE settings tested were the following: region size 2 with clip levels 2, 4 and 16; region size 4, with clip levels 2, 4 and 16; and region size 32 with clip levels 2, 4 and 16. The default or "unprocessed" settings correspond to the background image not undergoing CLAHE processing at all, which is equivalent to CLAHE processing with a clip of 0 and a region size of 512 (i.e. a single region covering the entire background). There were thus a total of 10 CLAHE settings tested in this experiment.

The digital images were printed onto standard 14X17 inch single emulsion film (3M HNC Laser Film, 3M, St Paul, MN) using a Lumisys Lumicam film printer (Lumisys Inc, Sunnyvale, CA). Each original 50 micron pixel was printed at a spot size of 160 microns, which produced film images 4x4 centimeters, resulting in an enlargement by a factor of three. The radiologist observers in the pilot experiment reported that the magnification did not make the backgrounds unrealistic. Forty images were printed per sheet of film. The images were randomly ordered and printed into thirty-two 8X5 grids on film. Both the film digitizer and film printer were calibrated, and measurements of the relationship between optical density on film and digital units on the computer were determined in order to generate transfer functions describing the digitizer and film printer. In order to maintain a linear relationship between the optical densities on the original analogue film and the digitally printed film, we calculated a standardization function that provided a linear matching between the digital and printer transfer functions. This standardization function was applied when printing the films to maintain consistency between the original optical

densities of the original mammography film and those reproduced on the digitally printed films. The film printer produces films with a constant relationship between an optical density range of 3.35 OD to 0.13 OD, corresponding to a digital input range of 0 to 4095, respectively.

There were 20 observers for the experiment. They were medical students and graduate students from the biomedical engineering and computer science departments. Performance bonus pay was provided. Observers selected the orientation of the spiculation within the image. All images contained a simulated spiculation in one of four orientations, for a four alternative-forced choice design. Observers were instructed to make their best guess if they could not see the spiculation or determine its orientation in a particular image.

Films were displayed in a dark room on a standard mammography viewbox that was masked to exclude excess light. Observers could move closer to the image, and could use a magnifying glass, if desired. A standard script was read to each observer prior to their participation, describing the goals of the research and the role of the observers in the study. Before actually starting the experiment, the observers were trained for the task through the use of three sets of images, including images in which the simulated object was very easy to detect. Thus the observers were quite familiar with the object that they were attempting to detect.

The order of presentation of stimuli was counterbalanced so as to eliminate any effects of learning and fatigue. All 160 possible combinations of processing conditions (10 CLAHE combinations of region size and clip level), contrast level (4 contrasts) and orientation of the simulated spiculations (4 orientations) were used in the experiment. The experiment was divided into 4 blocks, in which all 160 combinations appeared. Each observer saw all combinations in each block. All observers completed the experiment. There were 40 backgrounds. In each block, the 40 backgrounds are each paired with 160 possible processing condition combinations. The assignment was different for each block. Each observer examined 1280 images, for a total of 25600 total observations across all observers in the experiment. Each observer was assigned a different randomization of film order for the purpose of counterbalancing.

The experimental design can be thought of as a 3X3 factorial plus one additional condition. The factorial involves 3 clip levels (2, 4,16) crossed with 3 region sizes (2,8,32). In each of the 9 conditions in the factorial, the observer made 32 decisions at each of 4 contrast levels (10,25,40,55). In addition, each observer made 32 judgments at each of the 4 contrast levels with unenhanced images (clip=0, region=0). Therefore, each observer judged 3X3X4X32 plus 1X4X32 decisions, for a total of 1280 observations.

A total of 40 distinct background images from dense mammograms were used to create the stimuli. A phantom feature, the simulated spiculation, was added into the background. The image was then processed with CLAHE to yield the test stimulus. Each image was

used in each of 4 orientations to create 160 distinct backgrounds. Each background was used five times in a random order. Of the 32 decisions within each clip-region-contrast combination, 8 were made at each of 4 distinct spiculation orientations. (Table 1)

Observers took breaks after each block of images, and more often if necessary. No time limit was imposed on the observation of the images. Typically, the experiment took no more than 4 hours for each observer, divided into two sessions of 2 hours each. The two sessions were always scheduled on two different days within a week of each other.

Data Analysis Overview

Probit models were fit for each subject and enhancement condition using log10 contrast as the predictor. The probability that a subject gets a correct answer is assumed to be given by the following equation.

$$\Pr\{\text{correct}\} = 1/4 + (1 - 1/4)\ \Phi\ [(x - \mu_{ij})\sigma_i^{-1}]$$

where i indexes subject, and j indexes CLAHE settings. Here Φ indicates the cumulative Gaussian distribution function. For each subject, this gave a separate location parameter estimate for each CLAHE setting, and a common spread parameter estimate. Assuming a common spread parameter makes sense biologically, as it corresponds to an equal change in log contrast producing an equal change in perception, throughout the visual range. Also, the 1/4 arises from the 4 choice task.

The location parameter, μ_{ij} , is the mean of the corresponding Gaussian distribution for the ith subject and jth CLAHE setting. Processing conditions that improve detection performance will cause this parameter to be smaller, and the curve will shift to the left. This occurs because lower contrast levels are required to spot the object. When the processing of the image makes detection harder, higher contrast levels are needed to determine the orientation of the spiculation, and the curve shifts to the right. The values of σ_i , the spread parameter for the ith subject correspond to the slope of the curve. Larger values of σ_i correspond to steep slopes, or greater increase in detection rates per log contrast.

To compare the processing conditions and to examine the effect of window width and level, further analysis was needed. We defined an overall measure to be $\theta_{ij} = \mu_{ij} + \sigma_i$, which corresponds to the log contrast level at which the ith subject viewing the jth CLAHE condition scored 88% correct. We measured the "success" of a processing condition by calculating the difference between the θ score for the unprocessed image and the θ score for the condition for each subject, say $\delta j = \theta u - \theta j$, where u is unprocessed. A large positive δj score reflects improved performance. It indicates better detection with processed images than with unprocessed images.

Two analyses were performed using this outcome measure. To keep an overall nominal experiment-wise type 1 error rate of .05, a repeated measures analysis of variance was done at the .04 level, with a set of 9 T-tests at a .01/9 nominal level for each, and hence a

.01 level for the whole set.

Repeated measures analysis of variance (ANOVA) allows one to examine the effect of processing conditions and the interactions between region size and clip level, while accounting for the dependence of measurements taken on the same observer. The Geisser-Greenhouse corrected test was used throughout. The repeated measures ANOVA model was fitted, with the δj scores as the outcome. The log2 transformation of region size and clip level (log2reg and log2clip) are the predictors in this model.

Results

The repeated measures analysis of variance revealed that the interaction between region size and clip level was significant at the .04 level (p-value =0.0004, G-GΣ= 0.6987). Hence a series of (planned) step-down tests was implemented to investigate the nature of the interaction. The test of a linear-by-linear interaction was significant (p-value = 0.0002). (Figure 3)

At the nominal level of .01/9 = .0011, the differences between the default unprocessed condition and the CLAHE conditions were examined. Three settings of CLAHE processing conditions made finding the spiculations significantly easier and six made no significant difference. The settings that made detection easier were region size 32, with clip levels 2 and 4. There was one setting that significantly worsened detection

performance (region size 2 with clip level 16). (Table 2)

Average μ_{ij} and σ_i parameters from the best processing condition and the unprocessed condition were used to calculate a typical probit curve. Of the parameter values tested, the greatest improvement occurred for CLAHE processing with settings of region size=32 and clip level=2 (log2reg=5, log2clip=1). These values increased the correct detection of spiculations by 9 percent. This is shown in Figure 4.

Discussion

These results suggest that Contrast Limited Adaptive Histogram Equalization (CLAHE) can improve the detection of spiculations on dense mammographic backgrounds, if used properly. Our results also indicate that significant lesion visibility degradation can occur if the region size and clip levels are not chosen carefully. We believe that it is important to select the parameters to be applied in the testing of this tool in the clinic based on these types of careful analyses of laboratory studies. Preset parameter values might then be selected to apply to printed digital mammograms or to mammographic work stations where radiologists might interpret images "on line". Many radiologists who view CLAHE-enhanced mammograms have commented on the unpleasantness of the "image noise" that is rendered more visible when this algorithm is applied, and how it might cause worsening of their clinical performance. Our laboratory results support those concerns. If chosen poorly, CLAHE can degrade performance.

This work may not predict how this tool will function in a clinical setting. Specifically, graduate student observers and the use of simulated lesions might incorrectly predict the performance of radiologists in detecting real spiculations associated with real masses in real patients. We have demonstrated previously that graduate student performance at this task parallels the performance of experienced mammographers. (4) The signal to noise ratio and the type of image noise present in digital images might vary substantially from digitized mammograms when real full field digital images are used as the stimuli. Because we have used real clinical images and we have simulated lesions using relatively realistic stimuli, we are optimistic that this image processing algorithm will improve clinical performance. If so, radiologists might use CLAHE in the clinic as an adjunct to screening mammography whenever a mass is detected, much the way compression magnification views are used now. If the border characteristics, including the detection of subtle spiculation, is improved, radiologists might use this type of image processing to decide which lesions require further work-up.

Digital mammography is already available in a number of clinics in the US and Canada, including our own. It is highly likely that radiologists will want to apply image processing in an attempt to improve their performance in interpreting mammograms. The work reported here is intended to help radiologists narrow their choices regarding what might be clinically helpful before expensive clinical tests are undertaken. This project was intended to be a more rigorous exploration of the CLAHE parameters that might be used clinically in the most challenging areas in the breast, namely the dense parts.

This experiment does not address how CLAHE would affect the appearance of fatty areas of the breast, and the detection of spiculations in those parts. We would not want to view a mammogram solely with an algorithm applied that degrades performance in areas where sensitivity is currently quite high. By enhancing the visibility of image noise in fatty areas of the breast, CLAHE might degrade performance in these areas. It is possible that with effective training, radiologists might become used to improved visibility of background structures so that performance would not be degraded. However, if this algorithm is ultimately useful in dense areas only, it could potentially be applied selectively to only the dense parts of the breast. This could be accomplished by automatically segmenting the image to select for the densest parts and applying CLAHE only to those parts where it might provide benefit. Alternatively, it could be used as an adjunct with the image viewed in a standard format, and then with CLAHE applied to selected areas. In fact, we believe that CLAHE might be useful in this setting because it enhances the visibility of structures that extend across pixel boundaries, an apt description for the type of linear structure that a spiculation represents. Our results do not give us information about the performance of this algorithm in purely fatty areas of the breast, but the backgrounds used were relatively inhomogeneous in density, just as normal breast tissue is, and we expect these results to hold for all areas of the breast containing any soft-tissue density.

Our experiments to date cannot estimate the frequency of false positives when CLAHE would be used clinically. As discussed in our previous papers that explored the same

issues, alternate forced choice tests yield proportion correct as the primary outcome. Methods for converting proportion correct in this setting to a value for d', the sensitivity parameter of an ROC analysis, have been developed by Macmillan and Creelman (22). With this study design, and with the types of subjects and the amount of training used in this experiment, we believe that superior proportion correct will translate into superior d'. Of course, this must be proven in a true clinical setting with ROC analysis before these methods can be embraced for clinical purposes by practicing radiologists.

Legends

Figure 1

An example of a simulated spiculation used in the experiment.

Figure 2a

A specimen radiograph of a carcinoma showing spiculations. (arrows)

Figure 2b

The same carcinoma with a pseudospiculation inserted adjacent to the real spiculations (arrows)in the image. Note the extra linear structure running parallel to the 3 linear structures seen in figure 2a.

Figure 3

Interpolated predicted values from repeated measures ANOVA: difference in θ value versus region size and clip level. The peak shows the improved performance due to region size 32 with clip level 2.

Figure 4

Estimated detection probability for region size of 32 and clip level of 2. The shift in the curve to the left for the processed image reflects improved detection.

Table 1.

Table 1 displays the number of observations made per experimental condition by each observer.

Table 2.

Table 2 displays the difference in θ between images without and with CLAHE processing.

References

1. Homer MJ. Mammographic Interpretation: a practical approach. New York, NY, McGraw Hill, 1991, pp 4-5

2. Rosenman J, Roe CA, Cromartie R, et al. Portal Film enhancement: Technique and clinical utility. Int J Radiat Oncol Biol Physics 25(2): 333-338, 1993

3. Schmidt RA, Nishikawa RM. Clinical Use of Digital Mammography: The Present and the Prospects. JDI 8 (1, Suppl. 1): 74-79, 1995

4. Shtern F. Digital mammography and related technologies: a perspective from the National Cancer Institute. Radiology 183(3): 629-30, 1992

5. Feig SA, Yaffe MJ. Current Status of digital mammography. Seminars in US, CT and MR 17(5): 424-443, 1997

6. Pisano ED, Chandramouli J, Hemminger BM, et al. Does Intensity Windowing Improve the Detection of Simulated Calcifications in Dense Mammograms? J Dig Imaging 10(2): 79-84, 1997

7. Pisano ED, Chandramouli J, Hemminger BM, et al. The Effect of Intensity Windowing on the Detection of Simulated Masses Embedded in Dense Portions of Digitized Mammograms in a Laboratory Setting. J Dig Imaging. In press

8. Puff DT, Pisano ED, Muller KE, et al. A method for determination of optimal image enhancement for the detection of mammographic abnormalities. J Dig Imaging 7:161-171, 1994

9. McSweeney MB, Sprawls P, Egan RL. Enhanced Image Mammography. AJR 140:9-14, 1983

10. Smathers RL, Bush E, Drace J, et al. Mammographic microcalcifications: Detection with xerography, screen film, and digitized film display. Radiology 159: 673-677, 1986

11. Chan HP, Doi K, Galhorta S, et al. Image Feature analysis and computer-aided diagnosis in digital radiography: I. automated detection of microcalcifications in mammography. Med Phys. 14(4): 538-547, 1987

12. Chan HP, Vyborny CJ, MacMahon H, et al. Digital mammography ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. Investigative Radiology 22: 581-589, 1987

13. Hale DA, Cook JF, Baniqued Z, et al. Selective Digital Enhancement of Conventional Film Mammography. J Surg Onc 55: 42-46, 1994

14. Yin F, Giger ML, Vyborny CJ, et al. Comparison of Bilateral-Subtraction and Single-Image Processing Techniques in the Computerized Detection of Mammographic Masses. Investigative Radiology 28(6): 473-781, 1993

15. Yin F, Giger M, Doi K, et al. Computerized detection of masses in digital mammograms: Analysis of Bilateral Subtraction Images. Med Phys 18(5): 955-963, 1991

16. Kallergi M, Clarke LP, Qian W, et al. Interpretation of Calcifications in Screen/Film, Digitized and Wavelet-Enhanced Monitor-Displayed Mammograms: A Receiver-Operating Characteristic Study. Acad Radiology 3: 285-293, 1996

17. Pizer S, Zimmerman JB, Staab EV. Adaptive Grey Level Assignment in CT Scan Display. Journal of Computer Assisted Tomography 8(2): 300-305, 1984

18. Pizer SM. Psychovisual issues in the display of medical images. KH Hoehne, ed.,

Pictoral Information Systems in Medicine. Berlin, Springer-Verlag, 1985, pp 211-234

19. Revesz G, Kundel HL, Graber MD, The influence of structured noise on the detection of radiologic abnormalities. Investigative Radiology 9: 479-486, 1974

20. Kundel HL, Revesz G. Lesion conspicuity, structured noise and film reader error. AJR 126: 1233-1238, 1976

21. Revesz G, Kundel HL. Psychophysical studies of detection errors in chest radiology. Radiology 128: 559-562, 1977

22. MacMillan NA, Creelman CD. Detection theory: a user guide. Cambridge, England, Cambridge University Press, 1991, pp 135-136

Table 1.

Table 1. Number of Observations per Observer

REGION SIZE	CONTRAST	CLIP LEVEL			*TOTAL*
		2	4	16	
2	10	32	32	32	96
	25	32	32	32	96
	40	32	32	32	96
	55	32	32	32	96
8	10	32	32	32	96
	25	32	32	32	96
	40	32	32	32	96
	55	32	32	32	96
32	10	32	32	32	96
	25	32	32	32	96
	40	32	32	32	96
	55	32	32	32	96

REGION SIZE	CONTRAST	CLIP LEVEL	
		unenhanced	
unenhanced	10	32	32
	25	32	32
	40	32	32
	55	32	32
			1280

Table 2.

Table 2: Mean Difference Between CLAHE-processed and Unprocessed Theta Scores

Enhancement	Region Size	Clip Level	Difference Score	Standard Deviation	p-value
1	2	2	-0.002	0.044	0.8087
2	8	2	-0.007	0.047	0.5226
3	32	2	0.061	0.038	0.0001*
4	2	4	-0.019	0.045	0.0736
5	8	4	0.008	0.055	0.5076
6	32	4	0.053	0.045	0.0001*
7	2	16	-0.039	0.040	0.0004*
8	8	16	-0.036	0.058	0.0122
9	32	16	-0.031	0.062	0.0374

Note1: Larger difference scores correspond to better performance
Note2: A * indicates significance at the 0.0011 level

Average μij

ij and i parameters from the best processing condition and the unprocessed condition.

Appendix A.5

Pisano ED, Aylward S, Barbour P, Braeuning M.P, Brown ME, Chakraborty D, Cole E, Conant E, Eagle E, Fajardo LL, Feig S, Harrison J, Hemminger BM, R. Johnston RE, Jong R, Kennedy R, Kopans D, Kornguth P, Maidment A, Major S, McLelland R, Moore R, Muller K, Niklason L, Nishikawa R, Pizer SM, Plewes DB, Rosen E, Poyet C, Seaton K, Soo MS, Shumak R, Stahpit S, Staiger M, Vermont A, Walsh R, Williams MB, Williford M, Yaffe M, and Zong Z. Radiologist Preferences for Imaging Processing Algorithm for different clinical tasks for digital mammography display. Submitted to Radiology, Oct. 1999.

Radiologist Preferences for Digital Mammography Display

The International Digital Mammography Development Group

Etta D. Pisano, MD

Elodia B. Cole, MS

Stacey Major, MS

Shuquan Zong, MS

Bradley M. Hemminger, MS

Keith E. Muller, Ph.D.

R. Eugene Johnston, Ph.D

Ruth Walsh, MD

Emily Conant, MD

Laurie L. Fajardo, MD

Stephen A. Feig, MD

Robert M. Nishikawa, Ph.D.

Martin J. Yaffe, Ph.D.

Mark B. Williams, Ph.D.

Stephen R. Aylward, Ph.D.

Loren T. Niklason, Ph.D.

Andrew D.A. Maidment, Ph.D.

M. Patricia Braeuning, MD

Robert McLelland, MD

Andrei Vermont, MD

Phyllis Kornguth, MD, Ph.D.

Eric Rosen, MD

Mary Scott Soo, MD

Margaret Williford, MD

Daniel B. Kopans, MD

Dev Chakraborty, Ph.D.

Roberta Jong, MD

Rene Shumak, MD

Melinda Staiger, MD

Richard H. Moore, BA.

Stephen M. Pizer, Ph.D.

Donald B. Plewes, Ph.D.

Marylee E. Brown, BA

Etta Pisano, MD
Department of Radiology, University of North Carolina
CB#7510, 503 Old Infirmary, Chapel Hill, NC 27599-7510
(919) 966-1771, Fax (919) 966-0817

Elodia B. Cole, MS
Department of Biomedical Engineering, University of North Carolina
CB # 7575. 152 MacNider, Chapel Hill, NC 27599-7575
(919) 966-1148, Fax (919) 966-2859

Stacey Major, MS
Biostatistics Department, University of North Carolina
CB#7510, 503 Old Infirmary, Chapel Hill, NC 27599-7510
(919) 966-1771, Fax (919) 966-0817

Shuquan Zong, MS
Department of Biomedical Engineering, University of North Carolina
CB#7575, 152 MacNider, Chapel Hill, NC 27599-7575
(919) 966-1148, Fax (919) 966-2859

Bradley M. Hemminger, MS
Department of Radiology, University of North Carolina
CB# 7515, Chapel Hill, NC 27599-7515
(919) 966-2998, Fax (919) 966-2859

Keith E. Muller, Ph.D.
Biostatistics Department, University of North Carolina
CB# 7400, 3105C McGavran-Greenberg HL238
Chapel Hill, NC 27599-7400
(919) 966-7272, Fax (919) 966-3804

R. Eugene Johnston, Ph.D
Department of Radiology, University of North Carolina
Campus Box #7515, Chapel Hill, NC 27599-7515
(919) 966-5069, Fax (919) 966-2859

Ruth Walsh, MD
Department of Radiology, Duke University
Box 3808 Medical Center, Durham, NC 27710
(919) 684-7856, Fax (919) 684-7125

Emily Conant, MD
Breast Imaging Section, University of Pennsylvania Medical Center
3400 Spruce Street, Philadelphia, PA 19104-4238
(215) 662-4032, Fax (215) 662-7436

Laurie L. Fajardo, MD
Department of Radiology, Johns Hopkins University
601 N. Caroline St., BOX 0814 or Room 3140C
Baltimore, MD 21287
(410) 955-4794, Fax (410) 614-1079

Stephen A. Feig, MD
Division of Breast Imaging
Thomas Jefferson University Hospital
1100 Walnut Street - M.O.B./Ground Level
Philadelphia, PA 19107
(215) 955-7253, Fax (215) 923-7651

Robert M. Nishikawa, Ph.D.
Department of Radiology, University of Chicago
5841 South Maryland Ave - MC 2026
Chicago, IL 60637
(773) 702-9047, Fax (773) 702-0371

Martin J. Yaffe, Ph.D.
Department of Medical Biophysics and Medical Imaging, University of Toronto
Sunnybrook Health Sciences
2075 Bayview Avenue
North York, Ontario
M4N 35M, CANADA
(416) 480-6100 Ext. 5715, Fax (416) 480-5714

Mark B. Williams, Ph.D.
Department of Radiology, University of Virginia
Health Sciences Center, MR4 Building, Room 1174
Lane Road, Charlottesville, VA 22908
(804) 982-4422, Fax (804) 0924-9435

Stephen R. Aylward, Ph.D.
Department of Radiology, University of North Carolina
CB#7515, Chapel Hill, NC 27599-7515
(919) 966-9695, Fax (919) 962-1799

Loren T. Niklason, Ph.D.
3301 Carriage Trail, Hillsborough, NC 27278
(919) 732-8937, Fax (919) 732-6782

Andrew D.A. Maidment, Ph.D.
Department of Radiology, Thomas Jefferson University Hospital
1100 Walnut Street - M.O.B./Ground Level
Philadelphia, PA 19107
(215) 955-5013, Fax (215) 923-0268

M. Patricia Braeuning, MD
Department of Radiology, University of North Carolina
CB#7510, 503 Old Infirmary, Chapel Hill, NC 27599-7510
(919) 966-1771, Fax (919) 966-0817

Robert McLelland, MD
Department of Radiology, University of North Carolina
CB#7510, 503 Old Infirmary, Chapel Hill, NC 27599-7510
(919) 966-1771, Fax (919) 966-0817

Andrei Vermont, MD
Department of Radiology, University of North Carolina
CB#7510, 503 Old Infirmary, Chapel Hill, NC 27599-7510
(919) 966-1771, Fax (919) 966-0817

Phyllis Kornguth, MD, Ph.D.
Department of Radiology, Duke University
Box 3808 Medical Center, Durham, NC 27710
(919) 684-7655, Fax (919) 684-7125

Eric Rosen, MD
Department of Radiology, Duke University
Box 3808 Medical Center, Durham, NC 27710
(919) 684-2711, Fax (919) 681-8952

Mary Scott Soo, MD
Department of Radiology, Duke University
Box 3808 Medical Center, Durham, NC 27710
(919) 684-2711, Fax (919) 681-8952

Margaret Williford, MD
Department of Radiology, Duke University
Box 38038Medical Center, Durham, NC 27710
(919) 684-7644, Fax (919) 681-8952

Daniel B. Kopans, MD
Department of Radiology, Massachusetts General Hospital
Radiology, Level 2, Suite 219
15 Parkman Street, Boston, MA 02114
(617) 726-3093, Fax (617) 726-1074

Dev Chakraborty, Ph.D.
Department of Radiology, University of Pennsylvania
Physics Section, 3400 Spruce Street
Philadelphia, PA 19104
(215) 662-3268, Fax (215) 662-7011

Roberta Jong, MD
Department of Medical Imaging, Mount Sinai Hospital
600 University Avenue
Toronto, Ontario, M5G 1X5, Canada
(416) 586-5279, Fax (416) 586-4714

Rene Shumak, MD
Department of Radiology, University of Toronto
Sunnybrook Health Science Centre
2075 Bayview Avenue
Toronto, Ontario M4N 3M5 CANADA
(416) 480-4355, Fax (416) 480-5855

Melinda Staiger, MD
Department of Radiology, Good Samaritan Hospital Medical Center
Breast Health Center, 1000 Montauk Highway
West Islap, NY 11795
(516) 376-3624, Fax (516) 376-3655

Richard H. Moore, BA.
Massachusetts General Hospital
Level 2, Suite 219, 15 Parkman Street
Boston, MA 02114
(617) 726-3093, Fax (617) 726-1074

Stephen M. Pizer, Ph.D.
Biostatistics Department, University of North Carolina
CB# 7400, 3105C McGavran-Greenberg HL238
Chapel Hill, NC 27599-7400
(919) 966-7272, Fax (919) 966-3804

Donald B. Plewes, Ph.D.
Department of Medical Biophysics and Medical Imaging, Sunnybrook Health Sciences
Reichman Research Building, 2075 Bayview Ave - 6th FL
Toronto Ontario M4N 3M5 CANADA
(416) 480-5709, Fax (416) 480-5714

Marylee E. Brown, BA
Department of Radiology, University of North Carolina
CB# 7515, Chapel Hill, NC 27599-7515
(919) 966-7263, Fax (919) 966-2859

Corresponding Author: Etta D. Pisano, MD, Associate Professor of Radiology, 507 Old

Infirmary, CB# 7510, Chapel Hill, NC 27599-7510, phone: 919-966-6957, fax: 919-966-

0817, e-mail: etpisano@med.unc.edu

The following grants partially supported this work:

Grey Scale Image Processing for Digital Mammography, National Cancer Institute, National Institutes of Health, Grant # RO1 CA60193-05, Principal Investigator: Etta D. Pisano, University of North Carolina.

Multi-center Clinical Evaluation of Digital Mammography, Office of Women's Health, Department of Health and Human Services, Grant # 282-97-0078, Principal Investigator: Etta D. Pisano, University of North Carolina.

Evaluation of Digital Mammography Display, U.S. Army Medical Research and Material Command, Grant # DAMD 17-94-J-4345, Principal Investigator: Etta D. Pisano, University of North Carolina.

Enhancement and Analysis of Digital Mammograms. Canadian Breast Cancer Research Initiative, Grant #7289, Principal Investigator: Martin Yaffe, University of Toronto.

Clinical Evaluation of digital Mammography. National Cancer Institute, National Institutes of Health. Grant # 1RO1CA60192. Principal Investigator: Stephen A. Feig, Thomas Jefferson University.

Measurement of Differential Image Quality to investigate two methods for evaluating differential image quality due To compression algorithms.
National Cancer Institute, National Institutes of Health. Grant # RO1-CA75145-01A1.
Principal Investigator: Dev Chakraborty, University of Pennsylvania.

Computer Analysis of Mammography Phantom Images (CAMPI): an Application to the Optimization and Evaluation of a Full-field Digital Mammographic Machine To apply two powerful image quality evaluation tools to a full field digital mammography machine.
Office of Women's Health, Department of Health and Human Services, Grant # 282-97-0077. Principal Investigator: Dev Chakraborty, University of Pennsylvania.

Computer Aided Diagnosis in Digital Mammograms. National Cancer Institute, National Institutes of Health. Grant # RO1-CA60187. Principal Investigator: Robert Nishikawa, University of Chicago.

Radiologist Preferences for Digital Mammography Display

Author's Contributions:

Guarantor of Integrity of Entire Study: EDP

Study Concepts: EDP, BMH, REJ, MY, RMN

Study Design: EDP, BMH, KM, REJ, MY, RMN

Definition of Intellectual Content: All coauthors.

Literature Research: EDP

Clinical Studies: Not applicable.

Experimental Studies: EDP, EBC, BMH, MW, SA,SZ,LN, AM, MPB, RM, AV, PK, ER, MSS, RW, MW, EC, LLF, DK, and DC.

Data Acquisition: MPB, RM, AV, PK, ER, MSS, RW, MW, EBC, SZ

Data Analysis: SM, KM, EDP

Statistical Analysis: SM, KM.

Manuscript Preparation: EDP

Manuscript Editing: All coauthors.

Manuscript Review: All coauthors.

Abstract

Purpose: To determine the preferences of radiologists among eight different image processing algorithms applied to digital mammograms for the screening and diagnostic imaging tasks.

Materials and Methods: Twenty-eight images, representing pathologically proven cases obtained using three clinically available digital mammography units were processed and printed to film using Manual Intensity Windowing, Histogram-based Intensity Windowing, Mixture Model Intensity Windowing, Peripheral Equalization, MUSICA, Contrast Limited Adaptive Histogram Equalization, Trex® processing and Unsharp masking. Twelve radiologist observers compared the utility of the processed digital images to the screen-film mammograms of the same patient for breast cancer screening and breast lesion diagnosis.

Results: For the screening task, screen-film mammograms were preferred to all digital presentations, but Trex and MUSICA processed images were not statistically different in acceptability. All printed digital images were preferred to screen-film radiographs for the diagnosis of masses with Unsharp Masking processed mammograms statistically significantly preferred. For the diagnosis of calcifications, no digital algorithm was preferred to screen-film mammograms.

Conclusions: Radiologists prefer different digital processing algorithms for each of three mammography reading tasks, and for different lesion types. Softcopy display will eventually allow this option more easily.

Key Words: Digital mammography, image processing, display

Introduction

Aim of this Study

This study was performed to determine the preferences of radiologists regarding the display of printed digital mammograms. Specifically, eight different image processing algorithms were evaluated with respect to their utility for breast lesion characterization and breast cancer screening.

Background and Significance

Digital mammograms can be printed to film or displayed on a monitor. Typically, laser-printed films can display 4000X5000 pixels at 12 bits of grey scale. Although currently most radiologists are more comfortable with these printed images, the disadvantages of film display for digital mammography are obvious. Once an image is printed, it can no longer be manipulated, and any information available in the digital data but not captured in the printed image will therefore be lost.

With currently available high luminance, high resolution monitors (2000X2500 pixels) (1), only a portion of the breast can be displayed at one time at full resolution. In addition, comparing prior with current and left with right images is difficult. Roaming, zooming and grey level manipulation of the digital images with the computer, while possible, is not trivial to learn, and can be inefficient and time-consuming. Memory requirements for on-line interpretation are currently prohibitive. More practical displays with short, clinically acceptable display times for the entire set of images, including comparison images, are needed before digital mammography can reach its full

potential. Exploration of this issue was the purpose of a recent working group meeting jointly sponsored by the Office of Women's Health and the National Cancer Institute. (2)

Given the present limitations of soft-copy technology and radiologist preferences, digital mammography will most likely be displayed on film for the next few years at least. Therefore, exactly how the images should be printed is an important issue. Even if softcopy display is utilized, it is important to determine how the images should be viewed for optimal visualization of different lesion types in breasts of different radiographic density.

This is the first study to systematically explore the utility of displaying printed digital mammograms using 8 different image processing algorithms. We sought to determine the preferences of radiologists for algorithms for the two main tasks in mammography: lesion detection (screening) and characterization (diagnosis).

Methods

Image Production

Radiologist investigators at four participating institutions (EDP, LLF, DK and EC) selected digital mammograms for inclusion in the study. Studies were deemed eligible for inclusion if there were mammographic findings present and the screen-film image of the same patient was available for comparison. The cases were obtained using three different full field digital mammography devices: 10 cases from the Trex Digital Mammography System (Trex Medical Corporation, Long Island, NY), 10 cases from the Fischer SenoScan (Fischer Imaging Corporation, Denver, CO), and 8 cases from the General Electric Senographe 2000 D (General Electric Medical Systems,

Milwaukee, WI). Study cases consisted of standard unilateral mammograms containing mammographic findings.

The raw digital data was transmitted to the University of North Carolina, and to other participating institutions for image processing purposes, by Exabyte 8mm tape (Exabyte Corporation, Boulder, CO), or over the internet using File Transfer Protocols (FTP). Exabyte tapes were read using an Exabyte 8mm Tape Drive (Exabyte Corporation, Boulder, CO).

For Trex images, the image size was 4800x6400 pixels with 40 micron pixel size. For GE images, the image size was 1800x2304 pixels with 100 micron pixel size. For Fischer images, the image size was 3072x4800 pixels with 50 micron pixel size. All three units produce images with 16 bits/pixel.

All images were processed using each of 8 different algorithms: Manual Intensity Windowing (MIW), Histogram-based Intensity Windowing (HIW), Mixture Model Intensity Windowing (MMIW), Contrast Limited Adaptive Histogram Equalization (CLAHE), MUSICA (Agfa®), Unsharp Masking(UM), Peripheral Equalization (PE) and Trex® processing. The details regarding how these algorithms were applied for this study are described in Appendix 1.

All images were maintained at their original contrast and spatial resolution during processing. HIW, MIW, MMIW, and Trex processed images were printed to film without subsequent contrast manipulation of any type. CLAHE, PE and UM images were manually intensity windowed by an experienced mammography technologist before printing. MUSICA images were intensity windowed over a fixed range (0-4095 grey values). A single Orwin Model 1654 high brightness (100ftL) monitor (Orwin

Associates, Inc., Amityville, NY), utilizing a Dome Md5Sun Display Card (Dome Imaging, Waltham, MA) and a Sun UltraSparc model 2200 computer (Sun Microsystems, San Jose, CA) was used for all manual intensity windowing. Both the monitor and display card have a display matrix size of 2048 x 2560 pixels.

All images except those with Trex processing were printed on Kodak Ektascan HN film (Eastman Kodak Company, Rochester, NY) using a Kodak 2180 EktaScan Laser Film Printer® (Eastman Kodak Company, Rochester, NY). This printer is capable of 12 bits/pixel. Images that contained a bit range wider than that of the printer were linearly remapped to the range of the printer. Images were bilinearly interpolated by the Kodak printer to its maximum spatial resolution, with a 50 micron pixel size and a matrix of 4096 x 5120, and printed by the Kodak printer at this resolution. The laser film was processed using a Konica Medical Film Processor QX-400 (Konica Medical Corporation, Norcross, GA).

Trex processed images were printed on Agfa Scopix LT-2B helium-neon film using an Agfa LR5200 film printer (Agfa Division of Bayer Corporation, Ridgefield, NJ). This printer is capable of 8 bits per pixel. The matrix size for this printer is 4776x5944 pixels, and it has a 40 micron pixel size. Films were processed using a Kodak RP-Xomat processor (Eastman Kodak Corporation, Rochester, NY).

Trex mammograms were cropped from 4800x6400 pixels to fit the printer matrix size. Fischer and GE images were scaled up using interpolation by factors of 1.35 and 3.5 respectively. All printers and monitors used in this study were calibrated to comply with the DICOM grey scale display function standard. (American College of Radiology, Reston, VA and National Electrical Manufacturers Association, Roslyn, VA). (3)

Preference Study

A total of 65 lesions were identified and circled on the two views of a single version of the digital printed image of the patient's digital mammogram. A written description of each of the circled lesions was also prepared. This description included histologic information about the lesion, if that was available. Other lesions were presumed to be benign by virtue of a minimum of one year of mammographic stability with no clinical findings.

Tables Ia, Ib and Ic give a complete description of the images included in this study. Each case rated had at least 1 and up to 6 lesions to evaluate. Cases included only pathologically proven lesions (2 GE, 5 Trex and 2 Fischer), only presumed benign lesions (3 GE and 5 Fischer) and both types of lesions (3 GE, 5 Trex and 3 Fischer).

Twelve radiologists, all Mammography Quality Standards Act (MQSA) qualified mammography interpreters, independently participated as readers in this study. Written instructions were provided to each radiologist prior to the study. Appropriate masking of the viewboxes was utilized throughout.

The 28 cases were presented to each reader in random order by a research assistant. The craniocaudal images of each patient were presented first, followed by the mediolateral oblique images. The 8 processed digital mammograms were presented randomly within each case to each reader. Readers were also provided with the corresponding screen-film mammogram on the same patient, the annotated printed digital mammogram of the same view (lesions circled and numbered), and the description of the histologic findings for each case. The radiologists hung the

annotated image on the top viewbox panel of a standard mammography lightbox (Mammography Illuminator, Two Tier Desktop, Picker International, Inc., Norcross, GA). The screen-film mammogram and one of the eight digital processed mammogram to be rated were hung on the lower viewbox panel. Radiologists were provided with and encouraged to use a magnifying glass.

First, radiologists were asked to rate the visibility and characterizability of each lesion on the processed digital image with respect to its depiction on the corresponding screen-film mammogram. Radiologists were instructed to use their expert judgement in determining which areas on the screen-film image corresponded to the lesions seen on the digital images, taking into account differences in positioning, compression, and other factors. Utilizing all relevant clinical information, readers were asked to consider whether the processed digital image allowed sufficient visualization and characterization of each lesion so that the correct diagnosis could be reached. Each lesion on the digital mammogram was rated on a 5-point scale as much better, better, the same, worse or much worse than its screen-film counterpart (+2,+1,0,-1 or –2, respectively) with respect to visibility and characterizability. No magnification films or spot radiographs were provided to the readers.

Next, the radiologists were asked to rate the digital processed image as much better, better, the same, worse or much worse than the corresponding screen-film image for the purpose of screening (+2,+1,0,-1 or –2, respectively). For this task, they were asked to consider whether the digital image allowed sufficient visualization of all relevant anatomic structures for effective breast cancer screening. They were instructed to disregard artifacts that occurred outside the borders of the breast in

making this judgment. Again, craniocaudal images were rated first, followed by mediolateral oblique images.

The radiologists completed the tasks in the order in which they were presented. To limit the effects of fatigue, short breaks (at least 5 minutes) were required after every 50 minutes of work. The radiologists also took additional breaks as needed. On average, the radiologists required 5 hours to evaluate all images.

A research assistant recorded the radiologist's ratings for each processed digital image, as well as any other comments the radiologist made about the cases and/or the digital processing algorithms. The research assistant then manually entered the data into a Microsoft® Excel spreadsheet (Microsoft Corporation, Redmond, WA).

In sum, a total of 8 processed images for each of the 28 cases, minus the 7 images that were excluded, were compared to screen-film images by 12 radiologists. The total number of images viewed per radiologist was 441 (8 algorithms x 28 cases x 2 views = 448, then subtract the 7 that are missing). The same images were re-used for the 28x2=56 screening scores. The cases contained a total of 65 lesions, 29 that were pathologically proven and 36 that were presumed benign. Since there was one score per breast view for each of the 65 lesions within each algorithm for the diagnostic task, and an additional score for each view for each algorithm for the screening task, the total number of scores requested per radiologist was 1439. A total of 17268 scores were requested from the 12 readers.

As some readers intentionally or accidentally failed to rate one or more lesions, the dataset was incomplete. Some of the missing values were incurred when a reader was unable to detect a lesion on either the screen-film mammogram or on the digital

image, and was therefore unable to rate it. Missing scores for lesions not visible on screen-film were assigned scores of +2. To avoid possible bias towards digital due to positioning differences, the two cases for which scores were resolved in this manner were excluded from the final analyses (although including them did not change results). Missing scores for lesions not visible on the digital image were assigned scores of –2. Cases so affected were retained in all analyses.

Statistical Methods

All primary and exploratory analyses were conducted separately within the three mammography machines.

The primary analysis focused on the data for the diagnostic task, and consisted of two parts. First, a mean for each processing method by lesion type combination was calculated by averaging over reader, case, breast view, and lesion. Lesion types considered were calcifications and masses; masses with calcifications were classified as masses. Each of these 16 means (8 processing methods by 2 lesion types) was tested as equal to zero, corresponding to a null hypothesis of no difference in radiologist preference between the printed digital image and the screen-film mammogram. Per the Bonferroni technique for multiple comparisons, each test was evaluated at $\alpha = .01/16 = .000625$, for an overall Type I error rate of .01 for this set of tests.

In the second part of the primary analysis, model assumptions were verified and the data were analyzed by the Analysis of Variance (ANOVA) technique. The design for this two-way factorial repeated measures ANOVA included lesion type, processing

method, and their interaction. The test of method by lesion type interaction was conducted first, followed by step-down tests of the simple main effect of processing method within each lesion type. Within each of the two lesion types, there are (8 choose 2)=28 pairwise comparisons among the digital processing methods, for a total of (28*2)=56 tests. Per the Bonferroni technique, each test was evaluated at the α=.04/56=.000714285 level, resulting in an overall Type I error rate of .04 for this set of tests. Note that the overall Type I error rate for the complete primary analysis *within each machine* is (.01 + .04) = .05.

The exploratory analysis of the screening task data mirrored the primary analysis. First, a mean for each processing method by lesion type combination was calculated by averaging over reader, case and breast view. Lesion types considered were again calcifications and masses; masses with calcifications were classified as masses. Each of these 16 means (8 processing methods by 2 lesion types) was tested as equal to zero, corresponding to a null hypothesis of no difference in radiologist preference between the printed digital image and the screen-film mammogram with respect to breast cancer screening. Per the Bonferroni technique for multiple comparisons, each test was evaluated at α =.01/16 =.000625, for an overall Type I error rate of .01 for this set of tests. However, as this analysis is exploratory, p-values must be interpreted as descriptive statistics only.

In the second part of the exploratory analysis, model assumptions were verified and the data were analyzed by the Analysis of Variance (ANOVA) technique. The design for this two-way factorial repeated measures ANOVA included lesion type, processing method, and their interaction. The test of method by lesion type interaction

was conducted first, followed by stepdown tests of the simple main effect of processing method within each lesion type. Within each of the two lesion types, there are (8 choose 2)=28 pairwise comparisons among the digital processing methods, for a total of (28*2)=56 tests. Per the Bonferroni multiple comparisons procedure, each test was evaluated at the α=.04/56=.000714285 level, resulting in an overall Type I error rate of .04 for this set of tests.

Finally, all method by lesion type means were centered by subtracting the overall mean score for that machine. Centered means were computed for both the primary and exploratory analyses. In order to discourage comparison of mean scores among the different mammography machines, only the centered means will be presented in the results section. However, note that all p-values presented pertain to tests of the uncentered data.

All statistical analyses were performed using SAS Software, Version 6.12. (SAS Institute, Cary, NC.)

Results

Primary Analysis: Diagnostic Mammography Scores

Tables II and III show radiologist ratings of the digital processing algorithms with respect to the screen-film mammogram for the diagnostic mammography tasks. Ratings are presented by machine type. For all Tables, Machine A is the Fischer SenoScan, Machine B is The General Electric Senographe 2000D, and Machine C is the Trex Digital Mammography System.

For each machine, there was a strongly statistically significant relationship between lesion type and image processing algorithm preference for the lesion characterization, or diagnostic mammography, task (p=0.0002 for Fischer, 0.0024 for GE and 0.0338 for Trex). That is to say, for each machine, radiologists preferred different algorithms for the mass characterization and calcification characterization tasks.

Fischer Results

For the diagnostic evaluation of masses (including masses with calcifications), all printed digital mammograms were preferred to the screen-film mammograms for all eight processing algorithms. Musica, Trex, PE, UM and CLAHE were significantly preferred at the α =.01/16 =.000625 level. The machine-centered means for these algorithms were 0.37, 0.35, 0.32, 0.43 and 0.40, respectively.

For the diagnostic evaluation of calcifications, three of the eight printed processed digital mammograms, Trex processing, HIW, and MMIW, were rated as slightly better or equivalent to the screen-film mammograms (0.15, 0.07 and 0.03 machine-centered means respectively). These differences did not reach statistical significance. The screen-film image was significantly favored over the MIW and PE processed digital images, with p<.000625 (.01/16). The machine-centered means for these algorithms were -0.39 and -0.69, respectively.

GE Results

For the mass diagnostic task, the UM processed digital image was slightly but not statistically significantly preferred to the screen-film image. The machine-centered mean score for UM was 0.18. The screen-film mammogram was statistically significantly preferred over the Trex processed image at the α =.01/16 =.000625 level; the machine-centered mean score for Trex was -0.27.

For the calcifications diagnostic task, the MIW, HIW, UM, MMIW processed images were all slightly preferred to the screen film image. However, no digital processing algorithm was statistically significantly preferred. The machine-centered means for MIW, HIW, UM and MMIW were 0.19, 0.34, 0.30 and 0.28, respectively. The screen-film mammogram was statistically significantly preferred over the PE processed image at the α =.01/16 =.000625 level; the machine-centered mean score for PE was -0.48.

Trex Results

For the mass diagnostic task, all processed digital images except MMIW were preferred to the film-screen mammogram, with the Trex and HIW images statistically significantly preferred at the α =.01/16 =.000625 level. Machine-centered means for Trex and HIW were 0.53 and 0.57, respectively. The film-screen mammogram was preferred to the MMIW image, but not significantly. The machine-centered mean for MMIW was 0.17.

For the diagnostic evaluation of calcifications, the screen-film radiograph was statistically significantly preferred over all eight processed digital images at the α=.01/16

=.000625 level. The machine-centered mean scores ranged from –0.23 for the Trex processing algorithm to –0.75 for the PE method. These results were statistically significant for all eight algorithms (p<.01/16 or .000625).

Secondary Analysis: Overall Screening Score

Tables IV and V show radiologist ratings of the digital processing algorithms with respect to the screen-film mammogram for the screening mammography tasks. Ratings are presented by machine type.

There was a relationship between lesion type and image processing algorithm preference for each machine for the lesion detection, or screening mammography, score (p=0.0169 for Fischer, 0.1025 for GE and 0.0165 for Trex). Since this is an exploratory analysis, p-values may only be interpreted as descriptive statistics, and not as tests of significance.

Fischer Results

For the detection of both masses and calcifications, only Trex processed-digital radiographs were preferred to screen-film mammograms, although they were not strongly preferred. Machine-centered means for Trex processing of Fischer images were 0.84 for masses and 1.0 for calcifications. The screen-film image was strongly preferred over the MMIW-processed images for both mass and calcification detection. Machine-centered means for MMIW were -0.5 and –1.0 for mass and calcification detection, respectively. The screen-film image was also strongly preferred over MIW,

PE and UM for the detection of calcifications (machine-centered means of -0.27, -0.37 and -0.16, respectively). All tests were assessed at the $\alpha = .01/16 = 0.000625$ level.

GE Results

For the detection of both masses and calcifications, the screen-film mammograms were preferred to the printed digital radiographs for all processing algorithms. For masses, the machine-centered mean scores ranged from 0.44 for the Musica algorithm down to -0.48 for Trex. For calcifications, the machine-centered means ranged from 0.38 for Musica down to -0.41 for PE. All p-values were less than .01/16=0.000625 except Musica and HIW for masses, and Musica and MIW for calcifications.

Trex Results

The Trex-processed digital radiograph for the detection of masses was the only processing method preferred to the screen-film mammogram, but it was not strongly preferred ($p>.000625$). The Trex machine-centered mean for mass detection was 0.91. The screen-film mammogram was preferred to all other processed digital images for the detection of masses; centered means ranged from 0.91 for Trex down to –0.64 for MMIW. The screen-film mammogram was preferred to all eight processed digital images for the detection of calcifications; centered means ranged from 0.39 for Musica to –0.64 for CLAHE. P-values were less than 01/16=.000625 for all algorithms except Trex and MUSICA for both lesion types.

Discussion

Our results strongly indicate that radiologists prefer different processed versions of the digital mammogram depending on the task, the lesion type and the machine type. This conclusion suggests that digital mammograms would best be displayed using monitor systems that allow flexibility and easy, quick access to different processed versions of the images. If soft-copy interpretation is to become clinically practicable, ergonomic issues regarding image display using monitor systems must be overcome.

Undoubtedly, habit and experience influenced the preference of radiologists for screen-film images over processed digital images in many cases. A prior preference study, that attempted to exactly match the appearance of the screen-film mammograms through manual intensity windowing, showed that radiologists preferred digital mammography to screen-film. (6) Of course, such matching might not allow the full benefits of digital mammography to be realized.

This study is limited by the fact that it was a preference study and not a quantitative measure of how well the radiologists performed. Radiologists gave their opinions on which images would improve their performance. Certainly they made educated guesses, but a performance study would have been better at determining how mammographic interpretation would be affected by image processing. This study is a good first step, however. A performance study would require many more cases and would have been too expensive and unwieldy if 8 algorithms were tested. This experiment allows us to run the next study as a performance study, with more cases and fewer algorithms to test.

In addition, this study is limited in that the diagnostic mammography task did not include available compression and magnification views. However, since this affected both modalities equally, it should not have significantly altered our results.

Clearly, the entire universe of image processing algorithms has not been tested. We chose those algorithms that were available to us that are in clinical use, or that we believed might have clinical utility, and about which we had expertise. Perhaps wavelets or one of its derivatives or an algorithm yet to be developed might have performed better than all of those tested and the screen-film mammogram for all three tasks. In addition, a combination of algorithms, such as would be available with a softcopy display system, might allow for even better diagnostic performance and might have been the most preferred by the radiologists.

In addition, since we included such a small number of cases and different lesions were imaged with each of the three systems, we believe that the direct comparison of the results achieved by the three machines is not reasonable at this time. That is to say, we believe that the mean scores that the radiologists gave the various units for the various tasks should NOT be directly compared. We believe that we cannot justify statements about how the three units compare for the diagnosis or detection of masses or calcifications based on this preliminary study alone. For example, clearly the algorithms that were tested did not allow optimal calcification characterization with the Trex digital images, and optimal calcification and mass detection with the GE digital images. We believe that these results reflect more on the limitations of the algorithms tested than on the detectors themselves.

In fact, these results strongly suggest that each digital mammography manufacturer should determine which algorithms to use for optimal digital mammography display for each mammographic task. These results will help in guiding those decisions. Clearly, some sort of objective performance measure (7,8), rather than an aesthetic assessment, should be used by the manufacturers in guiding the selection of image processing algorithms. We believe that image processing might significantly enhance the achievable accuracy of digital mammography. Conversely, choices based on producing digital mammograms that closely resemble film-screen radiographs might limit the results that can be achieved with this new technology.

Finally, we did not have enough power in this study to determine whether other factors, such as breast density, patient age, location of the lesion within the breast and other variables, would influence radiologist preferences regarding the algorithms. The role of these factors will have to be evaluated in future studies.

Appendix 1

Manual Intensity Windowing (MIW)

For MIW, an expert mammography technologist manually intensity windowed the digital mammograms on an Orwin Model 1654 high brightness (100ftL) monitor (Orwin Associates, Amityville, NY), utilizing a Dome Md5Sun Display Card (Dome Imaging, Waltham, MA) and a Sun UltraSparc model 2200 computer (Sun Microsystems, San Jose, California). Both the monitor and display card have a display matrix size of 2048x2560 pixels. The intensity windowing software was interactive, and the technologist could choose either a linear or asymmetric sigmoidal within-window intensity mapping curve shape.

Histogram-Based Intensity Windowing (HIW)

In HIW, the histogram for each individual mammogram in a study is automatically analyzed in terms of its peaks and troughs. All components of the breast tissue, such as the parenchyma, fatty areas and skin edge portions, are recognized from these histogram features. With this method, contrast over the selected range of values of breast tissue is enhanced via simple intensity windowing.

Mixture-Model Intensity Windowing (MMIW)

MMIW uses a combination of geometric (i.e., intensity gradient-magnitude ridge traversal) and statistical (i.e., Gaussian mixture modeling) techniques. This method isolates the radiographically dense component in each mammogram and based on

statistical characteristics of this isolated region sets the parameters of an asymmetric sigmoidal intensity mapping function.

Contrast Limited Adaptive Histogram Equalization (CLAHE)

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a variant of Adaptive Histogram Equalization (AHE). In AHE, the histogram is calculated for the contextual region of a pixel, and the transformation provides the pixel a new intensity which is proportional to its rank in the intensity histogram. It is designed to provide higher contrast for pixel intensities which occur more frequently and to provide a single displayed image in which contrasts in all parts of the range of recorded intensities can be sensitively perceived. CLAHE limits the contrast increase factor produced by AHE to a user-specified unit. The CLAHE parameter settings (clip 4, region size 32) used in this study were based on prior published experiments. (7)

MUSICA

MUSICA processing is a multiscale wavelet-based contrast enhancement technique developed by Agfa® (Agfa Division of Bayer Corporation, Ridgefield Park, NJ). It involves variable enhancement of various spatial scale components of the image, followed by additive reconstruction. MUSICA processing was performed on an Agfa image processing workstation. Three of its four image processing parameters, namely Edge Contrast, Latitude Reduction and Noise Reduction were turned off by setting their levels to 0. The parameter for MUSICA was set to a maximum level of 5.

Unsharp Masking (UM)

Unsharp Masking is a technique used for crispening edges. A signal proportional to the unsharp, or low-passed filtered (blurred), version of the image is subtracted from the original image to yield a "sharpened" resulting image. The final image is produced by combining the original image (50% weighting) and the high-pass images (50% weighting). In our experiment a region size of 600x600 pixels was used for the calculation of the low-pass image.

Peripheral Equalization (PE)

IN PE, thickness differences between the periphery of the breast and the center portions are smoothed out so that the range of intensity values is accessible within the same narrow portion of the density look-up table. The thickness of the breast is approximated by using a smoothed version of the mammogram with resolution of about 3mm. The perimeter of the breast is determined by a simple threshold applied to the smoothed image, and grown to a few millimeters outside the breast. Masking of pixels outside this area is applied to remove detector flat-fielding artifacts. The thickness effect is removed essentially by dividing the original image values by those in the smoothed image. The correction is only applied within 3 cm of the periphery of the breast, while areas within the center of the breast are left at their original values. A damping factor, which limits the magnitude of the correction, is applied to the pixels immediately adjacent to the edge of the breast to reduce ringing. (8).

Trex Processing

The Trex processing used in this study is the proprietary processing applied as part of the Trex full-field digital mammography system. The algorithm is a weighted unsharp masking based on histogram data.

Acknowledgments

The authors gratefully acknowledges the contributions of the following individuals in support of this work: Faina Shtern of the Office of Women's Health, Department of Health and Human Services, Michelle Picarro, Michael Tesic, Ruth Grafton, Pat Campbell and Morgan Nields of Fischer Imaging Corporation, Beale Opsahl-Ong of General Electric Corporate Research and Development, Lim Cheung and Richard Bird of Trex Medical Corporation, Art Haus, Theresa Bogucki and Jeff Byng of Eastman Kodak, Willem Van Reit , Charles Augello and John Landry of Agfa; Mary Brown, Jenny Harrison, Patty Barbour, Lisa Quamme, Kerrie Kurgat, Mark Kramer, Sheldon Earp, Anna Cleveland, Sanjay Sthapit, and Joseph K.T. Lee of the University of North Carolina, Wendy Kurutz of Good Samaritan Hospital; Denise McDonald of the University of Pennsylvania; Anoma Gunasekara of the University of Toronto; Lisa Sparacino of Thomas Jefferson University; Jayne Cormier of Massachusetts General Hospital; Claire Poyet and Karen Seaton of Wake Radiology, Raleigh, NC, Elizabeth Eagle and Rebecca Kennedy of Greensboro Radiology, and Mary Baldwin of the University of Virginia.

This work would not have been possible without the generous support of Fischer Imaging Corporation, General Electric Medical Systems, Trex Medical Corporation, Eastman Kodak Corporation and the Agfa Division of Bayer Corporation.

References

1. Weibrecht M, Spekowius G, Quadflieg P, Blume H. Image quality assessment of monochrome monitors for medical soft copy display. SPIE 1997; 3031:232-244.

2. Shtern F and Winfield D. (eds.) Report of the Working Group on Digital Mammography: Digital Displays and Workstation Design. March 9-10, 1998. Public Health Service Office of Women's Health and National Cancer Institute Document.

3. http://www.nema.org/standards/DICOM.HTM

4. Moore RH, Kopans DB, Niklason LT, et al. Initial Clinical Experience with Full-field Digital Mammography. [abstract] Radiology 1997; 205(P): 274.

5. Pisano ED, Zong S, Hemminger BM, et al. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. J of Dig Imaging 1998; 11(4):193-200.

6. Pisano ED, Chandramouli J, Hemminger BM, et al. Does intensity windowing improve the detection of simulated calcifications in dense mammograms? J of Dig Imaging 1997; 10(2):79-84.

7. Pizer SM. Psychovisual issues in the display of medical images. In: Hoehne KH, ed. Pictorial Information Systems In Medicine. Berlin: Springer-Verlag, 1985; 211-234,.

8. Byng JW, Critten JP, Yaffe MJ. Thickness equalization processing for mammographic images. Radiology 1997; 203:564-568.

Table Ia. Case Types: Pathologically Proven Lesions. N=29

Machine	# of cases	Masses		Calcifications		Architectural Distortions	
		Cancer	Noncancer	Cancer	Noncancer	Cancer	Noncancer
General Electric	5	4	2	1	1		
TREX	10	4	5	2	4		
Fischer	5	1	2	1	1		1

Table Ib. Case Types: Lesions presumed benign due to stability. N=36.

Machine	# of cases	Masses	Calcifications	Architectural Distortions
General Electric	6	7	5	1
TREX	5	3	3	1
Fischer	8	5	10	1

Table Ic. Case Types: Histologic diagnoses for Pathologically Proven Lesions

Cancers	**Masses**	**Calcs**	**AD^**
Infiltrating Ductal Carcinoma	6	0	0
Ductal Carcinoma in Situ	1	4	0
Infiltrating Lobular Carcinoma	2	0	0
Noncancers			
Atypical Ductal Hyperplasia	1	2	0
Fibrocystic Change	1	2	0
Cyst	3	0	0
Fibroadenoma	2	0	0
LCIS** and Atypical Lobular Hyperplasia	1	1	0
Intraductal Papilloma	1	0	0
Atrophy	0	1	0
Chronic Inflammation and Fibrosis	0	0	1

AD^ = Architectural Distortion

LCIS** = Lobular Carcinoma in Situ.

Table II. Radiologist Preference Scoring for Image Processing Algorithms applied to Printed Digital Mammograms Relative to Screen-film Mammograms for Mass Diagnosis

Algorithm	Machine A	Machine B	Machine C
MUSICA	+	+	+
Trex Processing	+	-	+
MIW	+	+	+
HIW	+	-	+
PE	+	-	+
UM	++	+	+
CLAHE	++	-	+
MMIW	+	-	+

++Mean score statistically significantly better than screen-film mammogram
p<.000625 or 0.01/16
+Mean score better than screen-film mammogram.
- Mean score worse than screen-film mammogram.
--Mean score statistically significantly worse than screen-film mammogram
p<.000625 or 0.01/16

Meaningful comparisons between machine types is not possible using these data.

Table III. Radiologist Preference Scoring for Image Processing Algorithms applied to Printed Digital Mammograms Relative to Screen-film Mammograms for Calcification Diagnosis

Algorithm	Machine A	Machine B	Machine C
MUSICA	--	-	--
Trex Processing	+	-	-
MIW	--	+	--
HIW	+	+	--
PE	--	--	--
UM	-	+	--
CLAHE	-	+	--
MMIW	+	+	--

++Mean score statistically significantly better than screen-film mammogram
p<.000625 or 0.01/16
+Mean score better than screen-film mammogram.
- Mean score worse than screen-film mammogram.
--Mean score statistically significantly worse than screen-film mammogram
p<.000625 or 0.01/16

Meaningful comparisons between machine types is not possible using these data.

Table IV. Radiologist Preference Scoring for Image Processing Algorithms applied to Printed Digital Mammograms Relative to Screen-film Mammograms for Mass Detection Task

Algorithm	Machine A	Machine B	Machine C
MUSICA	-	-	-
Trex Processing	+	--	+
MIW	-	--	--
HIW	-	-	--
PE	-	--	--
UM	-	--	--
CLAHE	-	--	--
MMIW	--	--	--

++Mean score much better than screen-film mammogram
p<.000625 or 0.01/16
+Mean score better than screen-film mammogram.
- Mean score worse than screen-film mammogram.
--Mean score much worse than screen-film mammogram
p<.000625 or 0.01/16

Meaningful comparisons between machine types is not possible using these data.

Table V. Radiologist Preference Scoring for Image Processing Algorithms applied to Printed Digital Mammograms Relative to Screen-film Mammograms for Calcification Detection Task.

Algorithm	Machine A	Machine B	Machine C
MUSICA	-	-	-
Trex Processing	+	--	-
MIW	--	-	--
HIW	-	--	--
PE	--	--	--
UM	--	--	--
CLAHE	-	--	--
MMIW	--	--	--

++Mean score much better than screen-film mammogram
p<.000625 or 0.01/16
+Centered mean score better than screen-film mammogram.
- Centered mean score worse than screen-film mammogram.
--Centered mean score much worse than screen-film mammogram
p<.000625 or 0.01/16

Meaningful comparisons between machine types is not possible using these data.

Figure Legends

Figure 1a.

Photographic magnification of a craniocaudal view of a screen-film mammogram (1a).

Figure 1b.

A photographic magnification of a digital mammogram of the same region of the same breast imaged with a General Electric Senographe 2000 D (1b,c and d), processed with HIW. The clustered calcifications (arrows) seen in these images were needle-localized and surgically proven to be atypical ductal hyperplasia. The automated windowing algorithms, (MMIW, HIW, CLAHE) and MIW, the algorithms that somewhat compromise visibility of the skin for greater contrast in dense areas, all scored better than film screen for calcification characterization for this case. The algorithms that are designed to improve contrast while maintaining skin visibility either were equivalent to film screen (TREX processing) or worse (Unsharp Masking, MUSICA, and PE). (Case provided by Daniel Kopans of Massachusetts General Hospital).

Figure 1c.

Same digital mammogram processed with CLAHE.

Figure 1d.

Same digital mammogram processed with MIW.

Figure 2a.

This photographic magnification of a Fischer SenoScan® craniocaudal digital mammogram, processed with Unsharp Masking, shows a moderately well circumscribed mass in the far lateral portion of the breast, just below the skin, (arrow), that proved to be a simple cyst by ultrasound-guided fine needle aspiration. Because of its location at the periphery of the breast, the lesion is not even visible on some of the algorithms that cause reduced visibility of subcutaneous detail to allow improved penetration and contrast for the densest areas (MIW, MMIW and HIW).

Figure 2b.

The same area of the digital mammogram processed with CLAHE.

Figure 3a.

This photographic magnification of the subareolar region of a screen-film mammogram reveals a partially circumscribed, partially obscured nonpalpable mass (arrows) that had been visible for over 1 year by mammography and was therefore presumed benign. (Case provided by Emily Conant, MD of the University of Pennsylvania.)

Figure 3b.

The General Electric Senographe 2000 D digital mammogram of the same lesion (arrows) displayed after processing with Unsharp Masking, an algorithm preferred by study radiologists for mass characterization with GE images. Note the improved border conspicuity over the screen-film image.

Figure 3c.

Digital mammogram of the same lesion (arrows) displayed using MIW processing.

Figure 3d.

Digital mammogram of the same lesion displayed using MUSICA processing.

Figure 4a.

This photographic magnification of a mediolateral oblique screen-film mammogram shows a spiculated mass in the axillary tail that proved by core biopsy and subsequent mastectomy to be infiltrating lobular carcinoma and lobular carcinoma in situ. (Case provided by the Mark Williams, pH of the University of Virginia and Laurie Fajardo, MD of Johns Hopkins University.)

Figure 4b.

Trex Digital Mammography System digital mammogram of the same lesion, processed with MUSICA. For this lesion, all digital images had higher mean scores than did the film screen mammogram, probably because the spiculations on the anterior margin of the mass are more obvious on the digital images. The five processed digital images, 4b-4f, are shown in their order of preference to the radiologists for this particular case.

Figure 4c.

Processed with CLAHE.

Figure 4d.

Processed with HIW.

Figure 4e.

Processed with MIW.

Figure 4f.

Processed with Trex processing.

Appendix A.6

Pisano ED, Aylward S, Barbour P, Braeuning M.P, Brown ME, Chakraborty D, Cole E, Conant E, Eagle E, Fajardo LL, Feig S, Harrison J, Hemminger BM, R. Johnston RE, Jong R, Kennedy R, Kopans D, Kornguth P, Maidment A, Major S, McLelland R, Moore R, Muller K, Niklason L, Nishikawa R, Pizer SM, Plewes DB, Rosen E, Poyet C, Seaton K, Soo MS, Shumak R, Stahpit S, Staiger M, Vermont A, Walsh R, Williams MB, Williford M, Yaffe M, and Zong Z. Image Processing Algorithms for Digital Mammography – A Pictorial Essay. Submitted to Radiographics, Oct. 1999.

Image Processing Algorithms For Digital Mammography -A Pictorial Essay

Etta D. Pisano, MD*

Elodia B. Cole*

Bradley M. Hemminger, MS*

Martin J. Yaffe, PhD~

R. Eugene Johnston, PhD*

Mark B. Williams, PhD†

Stephen R. Aylward, PhD*

Andrew D.A. Maidment, PhD‡

Loren T. Niklason, PhD

Emily F. Conant, MD²

Laurie L. Fajardo, MDϕ

Daniel B. Kopans, MD'

Marylee E. Brown*

Stephen M. Pizer, PhD*

*University of North Carolina at Chapel Hill
~University of Toronto
ϕJohns Hopkins University
†University of Virginia
²University of Pennsylvania
'Massachusetts General Hospital
‡Thomas Jefferson University

This work was supported in part by National Cancer Institute RO1 CA60193-05; Office of Women's Health, Department of Health and Human Services, # 282-97-0078; U.S. Army Medical Research and Material Command, # DAMD 17-94-J-4345, Canadian Breast Cancer Research Initiative #7289; National Cancer Institute 1RO1CA6019National Cancer Institute RO1-CA75145-01A1 and National Cancer Institute RO1-CA60183.

Corresponding Author: Etta D. Pisano, The University of North Carolina at Chapel Hill, 503 Old Infirmary, CB# 7510, Department of Radiology, Chapel Hill, NC 27599-7510, Phone: (919) 966-6957, Fax: (919) 966-0817, E-mail: etpisano@med.unc.edu

ABSTRACT

This article demonstrates the use of image processing algorithms with digital mammograms. Four illustrative cases obtained using three different digital mammography units show the advantages and disadvantages of seven different display algorithms for the specific tasks required in breast imaging – diagnosis and screening. This paper will elucidate why different algorithms may be useful for different tasks. The use of multiple algorithms for digital mammography display will necessitate the development of softcopy workstations for this modality.

Summary Statement

This article demonstrates the use of image processing algorithms with digital mammograms.

INTRODUCTION

The effectiveness of digital mammography in breast cancer detection is currently under investigation. This imaging modality separates image acquisition and image display, which allows for optimization of both.

In screen-film mammography, film serves as the medium for both image acquisition and display. Screen-film mammography offers limited detection capability of low contrast lesions in dense breasts. This poses a problem for the estimated 40% of women with dense breasts who receive mammograms (1). In this population, diagnosis often requires additional imaging, which results in more radiation exposure for the patient. When additional images fail to provide useful diagnostic information, a decision must be made as to whether the suspicious regions require biopsy or short or long term follow-up. Because of the expense and the risk associated with additional radiation exposure and surgery, any method of image presentation that increases the diagnostic conspicuity of lesions in breast tissue, but especially in dense tissue, would be a significant advance.

Digital mammography systems, unlike screen-film mammography systems, allow for manipulation of fine differences in image contrast through the use of image processing algorithms. As a result, very subtle differences between abnormal and normal but dense tissue can be made more obvious. The purpose of this paper is to illustrate the appearance of various image processing algorithms for display of digital mammograms

and to discuss how these algorithms may affect the ability of radiologists to interpret the images.

Cases Used in this Paper

The four cases used in this paper to demonstrate the image processing algorithms were selected to show the range of types of mammographic lesions and the potential advantages and disadvantages of the different display algorithms. Figures 1a, 2a, 3a and 4a show the screen-film radiographs of the same patients.

Figure 1a shows a photographic magnification of a partially obscured and partially circumscribed mass that proved to be a simple cyst by ultrasound and needle aspiration. The accompanying digital mammogram, displayed with 7 different image processing algorithms in Figures 1b-1h, was acquired at the University of North Carolina using the Fischer SenoScan full field digital mammography unit (Fischer Imaging Corporation, Denver, CO).

Figure 2a shows a screen-film mammogram of a breast with two indistinct masses. Photographic magnification of the screen-film mammogram of the larger mass is provided in Figure 2b. Both masses proved to be infiltrating ductal carcinoma with accompanying ductal carcinoma in situ (DCIS) by needle-localized open surgical biopsy. Figures 2c through 2f show the same patient's digital mammogram, which was acquired at Massachusetts General Hospital using the General Electric Senographe

2000D full field digital mammography system (General Electric Medical Systems, Schenectady, NY).

Figures 3a and 3b show a screen-film mammogram of a palpable spiculated mass that proved to be infiltrating ductal carcinoma with associated cribiform and solid-type DCIS at open surgical biopsy. Figures 3c and 3d show the Fischer SenoScan digital mammogram of the same patient, acquired at the University of North Carolina.

Figure 4a is a photographic magnification of a screen-film mammogram containing a pleomorphic cluster of calcifications that proved to be atrophic breast tissue at stereotactically-guided core biopsy. Figures 4b-4h show the same patient's digital mammogram from a Trex Digital Mammography System (Trex Medical Imaging Corporation, Danbury, CT), acquired at the University of Virginia.

Brief Overview of the Digital Mammography Systems

The GE system produces images with a spatial resolution of 100 microns per pixel that have a total matrix size of 1800 x 2304 pixels. Trex images are 41 microns per pixel with a display matrix size of 4800 x 6400 pixels. Fischer images are 54 microns per pixel with an image size of 3072 x 4800 pixels. The smaller the number of microns per pixel, the smaller the features that can be measured in the image produced. As for contrast resolution, the Trex and GE units offer 14 bits per pixel while the Fischer unit offers 12 bits per pixel. Increasing contrast gradation provides the opportunity to distinguish finer and finer density differences between features in the image. However,

the ability of a human observer to distinguish finer and finer gradations of gray may not always be possible due to visual perception and display device limitations. Detailed descriptions of the image acquisition hardware are provided elsewhere. (2)

Image Processing Algorithms Illustrated

Each manufacturer has developed image processing algorithms to use with its acquisition system. In addition, there are a number of algorithms that have been developed by independent investigators for use with digital mammograms. Specifically, the seven algorithms that are demonstrated in this paper are Manual Intensity Windowing (MIW), Histogram-based Intensity Windowing (HIW), Mixture-Model Intensity Windowing (MMIW), Contrast-Limited Adaptive Histogram Equalization (CLAHE), Unsharp Masking (UM), Peripheral Equalization (PE), and Trex proprietary processing.

Intensity Windowing Algorithms (IW)

Intensity windowing algorithms act on individual pixels within an image. A small portion of the full intensity range of an image is selected and then remapped to the full intensity range of the display device. This allows for the selection of specific intensity values of interest. For example, intensity values that represent abnormal tissue and dense but normal tissue are selected to allow for the exaggeration of small differences in intensity values between the two objects, thus potentially increasing the conspicuity of any

abnormal regions. The three versions of IW demonstrated in this paper are Manual Intensity Windowing (MIW), Histogram-based Intensity Windowing (HIW), and Mixture-Model Intensity Windowing (MMIW). These algorithms differ in how intensity values of interest are selected.

Manual Intensity Windowing (MIW)

Manual Intensity Windowing was performed by an expert mammography technologist who interactively adjusted the contrast levels as appropriate for each image using an Orwin 1654 high brightness monitor (Orwin Associates, Amityville, NY) and a Sun Ultra Sparc 2200 (Sun Microsystems, San Jose, CA). The goal of this algorithm is to manually reproduce the appearance of a screen-film mammogram.

Figures 1b, 2c and 4b all illustrate this algorithm applied to the selected cases. These images readily demonstrate how similar in appearance the digital mammograms can be to standard screen-film mammograms of the same patients. For Figure 2c, the center of the large mass is very light. This is because of the technologist's selection of a window that allowed visualization of both lesions in the image. Both lesions were obvious to her trained eyes. In order to keep the smaller lesion from appearing less obvious or even disappearing completely, she windowed the larger lesion so it was slightly lighter than ideal.

This case points out the obvious limitation of this interactive windowing algorithm. It is operator dependent. A less experienced operator might choose different windows that could obscure some of the visible pathology.

Histogram-based Intensity Windowing (HIW)

Histogram-based intensity windowing utilizes a histogram of intensity values of the digital image to *automatically* identify breast tissue areas of interest and applies a simple intensity window to these regions of interest. For example, the skin edge intensity values are low . The densest parts of the breast have high intensity areas. The computer automatically identifies the dense areas and windows the image depending on the amount of dense and fatty tissue. This allows for an individualized window based on each patient's histogram.

Figures 1c and 4c demonstrate this automated windowing algorithm. For the cyst in Figure 1c, notice the improved conspicuity of the lesion edge on the digital radiograph compared to the screen-film mammogram shown in Figure 1a. Part of the difference in visibility in the lesion border and the accompanying benign calcifications is attributable to differences in positioning and compression. There is some loss of detail outside the dense parts of this image compared to the screen-film image and to the other digital mammogram presentations. This might detract from the use of this algorithm for screening.

Mixture-Model Intensity Windowing (MMIW)

Mixture-Model Intensity Windowing segments the breast utilizing a combination of geometric (i.e., gradient magnitude ridge traversal) and statistical (i.e., Gaussian mixture modeling) techniques into fatty, mixed, and dense tissue regions. Only the radiographically densest portions of the mammogram are selected for image processing. Once the dense regions are defined, intensity windowing is applied to the region of interest.[3]

Figures 1d, 2d, 2e and 4d demonstrate digital mammograms with MMIW applied. For all three cases, this algorithm enhances the visibility of the lesion borders against the fatty background. However, the mixed parenchymal densities that abut the lesion are lost in some cases. This effect is most dramatic at the edges of the mammogram, as shown in Figure 2d. Clearly, if this type of statistical sampling of the image is utilized to determine an optimal intensity window, an additional algorithm that enhances the visibility of the periphery of the breast should be used to rescue information that is lost at the low density subcutaneous regions of the breast.

Both HIW and MMIW algorithms might be useful on a workstation. At the touch of a button, radiologists could request a processed digital mammogram that allows them to see through the densest portions of the breast. Neither would probably be acceptable for the display of screening mammograms, however, since information in the peripheral and fatty areas of the breast is not visible when these algorithms are applied.

Contrast Limited Adaptive Histogram Equalization (CLAHE)

Contrast Limited Adaptive Histogram Equalization is a special class of Adaptive Histogram Equalization (AHE). In AHE, the histogram is calculated for small regional areas of pixels, producing local histograms. These local histograms are then equalized or remapped from the often narrow range of intensity values indicative of a central pixel and its closest neighbors to the full range of intensity values available in the display. CLAHE limits the maximum contrast adjustment that can be made to any local histogram. The CLAHE parameter settings (clip 4, region size 32) used in these sample digital mammograms were selected based on previous experiments (4). After CLAHE was applied, Manual Intensity Windowing was used so that the contrast of the resulting image more closely approximated standard screen-film mammography.

Figures 1e and 4e demonstrate CLAHE-processed digital mammograms. The lesions in these images do appear very obvious compared to background and the image detail is very good. However, note also the obvious visualization of graininess in the digital images. This is due to the enhanced visibility of both image signal and image noise by this algorithm. Again, this algorithm might be helpful in allowing radiologists to see subtle edge information, such as spiculation. It might degrade performance in the screening setting by enhancing the visibility of nuisance information that could simulate calcifications.

Unsharp Masking (UM)

Unsharp masking (5) is a technique by which a low-pass filtered version of the original image is created and the image values that result are subsequently subtracted from the original image. The resultant high-pass image is then added to the original image, which produces the final image with accentuated edges. Manual Intensity windowing was then applied to the resultant image to adjust the contrast to levels more closely approximating standard screen-film mammography.

Figures 1f, 2f, 3c and 3d demonstrate UM applied to digital mammograms. The sharpness of the borders of the mass lesions is enhanced, as is the intended effect of this algorithm. The spiculations in the Fischer digital mammogram, seen in Figures 3c and 3d, are rendered especially evident. Of course, Figure 2f illustrates how even an indistinct mass can appear more circumscribed when this algorithm is applied, obviously an undesirable outcome if this were to lead to inappropriate patient follow-up instead of biopsy.

Peripheral Equalization (PE)

Peripheral Equalization is a technique that enhances the periphery of the breast. (6) There are variations in thickness of the breast tissue under compression during image acquisition. The outer edges of the breast, which are not as thick as the interior, are typically over-penetrated by x-rays at acquisition. This results in the periphery being

difficult to distinguish visibly from the black film background. In PE, a significantly less detailed version of the mammogram is used to approximate breast thickness. Thresholding is applied to the resulting image and extended a bit to determine the breast perimeter. This algorithm does not affect the interior regions of the breast. After PE was applied, Manual Intensity Windowing was used to adjust the resultant image contrast to more closely resemble a traditional screen-film mammogram.

Figures 1g and 4g demonstrate this image processing algorithm. Both calcification and mass details are well depicted in these images. In addition, as is especially evident in Figure 1g, the peripheral information in the surrounding breast is preserved. This algorithm might be effective in the screening setting since it preserves image features in all breast locations. However, there does appear to be some flattening of image contrast in the nonperipheral portions of the mammograms when this algorithm is applied.

Trex-Processing

Trex-processing was developed by Trex Medical Imaging Corporation for use with the Trex Digital Mammography System. This method utilizes a form of histogram-based unsharp masking.

This algorithm is demonstrated in Figures 4h and 4i. As can be seen from these images, the algorithm allows visualization of both lesion detail and breast edge information. This is achieved with some flattening of image contrast, however, as seen in this case when

the Trex-processed version is compared to the other processed versions of the same image.

Summary

It is obvious from the illustrated cases that different digital image processing algorithms are likely to be useful for different tasks. Characterization of lesions and screening will most probably require a uniquely adapted image-processing algorithm to provide the best presentation for visualization of different image features. In addition, different types of lesions, masses and calcifications, might benefit from specifically tailored algorithms. This will not be easily achieved unless the current method of displaying mammograms on film is replaced by a softcopy display system.

Given the added costs, the efficacy of digital mammography will ultimately depend upon improved diagnostic accuracy over conventional screen-film mammography. The development and assessment of image processing methods that allow for detection and characterization of individual lesion types will be instrumental in the acceptance of this new technology.

References

1. Shtern, F. Digital mammography and related technologies: a perspective from the National Cancer Institute. Radiology. 1992; 183: 629-630.

2. Feig SA, Yaffe MJ. Current Status of Digital Mammography. Seminars in US, CT and MRI. 1996; 17: 424-443.

3. Aylward SR, Hemminger BM, Pisano ED. Mixture modeling for digital mammogram display and analysis. In: Karssemeijer N, Thijssen M, Hendriks J, van Erning, eds. Digital Mammography Nijmegen, 1998. Dordrecht: Kluwer Academic Publishers, 1998; 305-312.

4. Pisano ED, Zong S, Hemminger BM, DeLuca M, Johnston RE, Muller K, Braeuning MP, Pizer S. Contrast Limited Adaptive Histogram Equalization Image Processing to Improve the Detection of Simulated Spiculations in Dense Mammograms. Journal of Digital Imaging. 1998; 11(4): 193-200.

5. Chan HP, Vyborny CJ, MacMahon H, et al. Digital mammography ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. Investigative Radiology. 1987; 22: 581-589.

6. Byng, J.W., Critten J.P. and Yaffe, M.J. Thickness equalization processing for mammographic images. Radiology 1997; 203:564-568.

Legends

Figure 1a: This is a photographic magnification of a craniocaudal screen-film mammogram of a cyst.

Figure 1b: Photographic magnification of the Fischer digital mammogram, processed with Manual Intensity Windowing (MIW).

Figure 1c: Photographic magnification of the Fischer digital mammogram, processed with Histogram-based Intensity Windowing (HIW).

Figure 1d: Photographic magnification of the Fischer digital mammogram, processed with Mixture-Model Intensity Windowing (MMIW), showing the same lesion as seen in Figure 1a.

Figure 1e: Photographic magnification of the Fischer digital mammogram, processed with Contrast Limited Adaptive Histogram Equalization (CLAHE).

Figure 1f: Photographic magnification of the Fischer digital mammogram, processed with Unsharp Masking (UM). (Algorithm provided by Andrew Maidment, PhD of Thomas Jefferson University).

Figure 1g: Photographic magnification of the Fischer digital mammogram, processed with Peripheral Equalization (PE). (Algorithm provided by Martin Yaffe, PhD and Gordon Mawdsley, PhD of the University of Toronto)

Figure 2a: This mediolateral oblique screen-film mammogram shows two masses, (arrows) which both proved to be infiltrating ductal carcinoma with associated ductal carcinoma in situ at open surgical biopsy. (Courtesy of Daniel Kopans, MD, of Massachusetts General Hospital.)

Figure 2b: Photographic magnification of the screen-film image of the large inferior carcinoma.

Figure 2c: A photographic magnification of the larger lesion seen on the digital mammogram, displayed with MIW.

Figure 2d: This General Electric digital mammogram, processed with Mixture-Model Intensity Windowing (MMIW), shows both cancers very well.

Figure 2e: A photographic magnification of the larger lesion seen on the digital mammogram, displayed with MMIW.

Figure 2f: A photographic magnification of the larger lesion seen on the digital mammogram, displayed with UM. (Algorithm provided by Andrew Maidment, PhD of Thomas Jefferson University.)

Original Photos Available Upon Request

Figure 3a: This mediolateral oblique screen-film mammogram shows a spiculated mass in the axillary portion of the breast, an infiltrating ductal carcinoma with associated cribiform and solid-type ductal carcinoma in situ at open surgical biopsy.

Figure 3b: A photographic magnification of the lesion seen on the screen-film mammogram.

Figure 3c: The Fischer digital mediolateral oblique mammogram, displayed using Unsharp Masking. (Algorithm provided by Andrew Maidment, PhD of Thomas Jefferson University.)

Figure 3d: A photographic magnification of the lesion seen on the digital mammogram, displayed with Unsharp Masking. (Algorithm provided by Andrew Maidment, PhD of Thomas Jefferson University.)

Figure 4a: This photographic magnification of a screen-film mammogram revealed a cluster of calcifications, which proved to be atrophic breast tissue at core biopsy. (Case provided by the University of Virginia and Laurie Fajardo of Johns Hopkins University.)

Figure 4b: The MIW processed digital mammogram with photographic magnification of the clustered calcifications.

Figure 4c: The HIW processed digital mammogram.

Figure 4d: The MMIW processed digital mammogram.

Figure 4e: The CLAHE-processed digital mammogram.

Figure 4f: The UM processed digital mammogram. (Algorithm provided by Andrew Maidment, PhD of Thomas Jefferson University.)

Figure 4g: The PE processed digital mammogram. (Algorithm provided by Martin Yaffe, PhD and Gordon Mawdsley, PhD, of the University of Toronto.)

Figure 4h: The digital mammogram with Trex proprietary processing applied.

Figure 4i: A photographic magnification of the lesion as seen with the Trex processing.